

“信息化与信息社会”系列丛书之

高等学校信息安全专业系列教材

信息隐藏概论

陆哲明 聂廷远 吉爱国 编著

電子工業出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

信息隐藏技术是一种重要的现代信息安全技术。本书全面介绍了信息隐藏的概念、发展现状、各个研究分支的基础理论和技术方法及其在信息安全领域的应用。本书首先介绍信息隐藏的基本概念、模型、研究分支、发展现状和相关理论与技术。其次,详细介绍三类主要的信息隐藏技术——隐写术、数字水印技术和数字指纹技术的概念、基础理论和主要方法。接着,介绍一种近年来比较热门的无损信息隐藏技术以及信息隐藏技术的其他研究分支。然后,分别介绍针对隐写技术和数字水印的攻击技术——隐写分析和数字水印攻击技术的基本概念、分类和典型方法。最后,介绍信息隐藏技术在知识产权保护、内容认证和保密通信等领域的典型应用。

本书可作为高等院校具有一定计算机基础的信息安全专业、电子信息工程专业、计算机专业、通信工程专业的研究生或高年级本科生的教材或参考书,也可作为科研院所相关专业的科技工作者的参考书。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有,侵权必究。

图书在版编目(CIP)数据

信息隐藏概论 / 陆哲明, 聂廷远, 吉爱国编著. —北京: 电子工业出版社, 2014.11

(信息化与信息社会系列丛书)

高等学校信息安全专业系列教材

ISBN 978-7-121-24390-5

I. ①信… II. ①陆… ②聂… ③吉… III. ①信息系统—安全技术—高等学校—教材 IV. ①TP309

中国版本图书馆 CIP 数据核字 (2014) 第 220491 号

策划编辑: 刘宪兰

责任编辑: 郝黎明

印 刷:

装 订:

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本: 787×1092 1/16 印张: 26.25 字数: 672 千字

版 次: 2014 年 11 月第 1 版

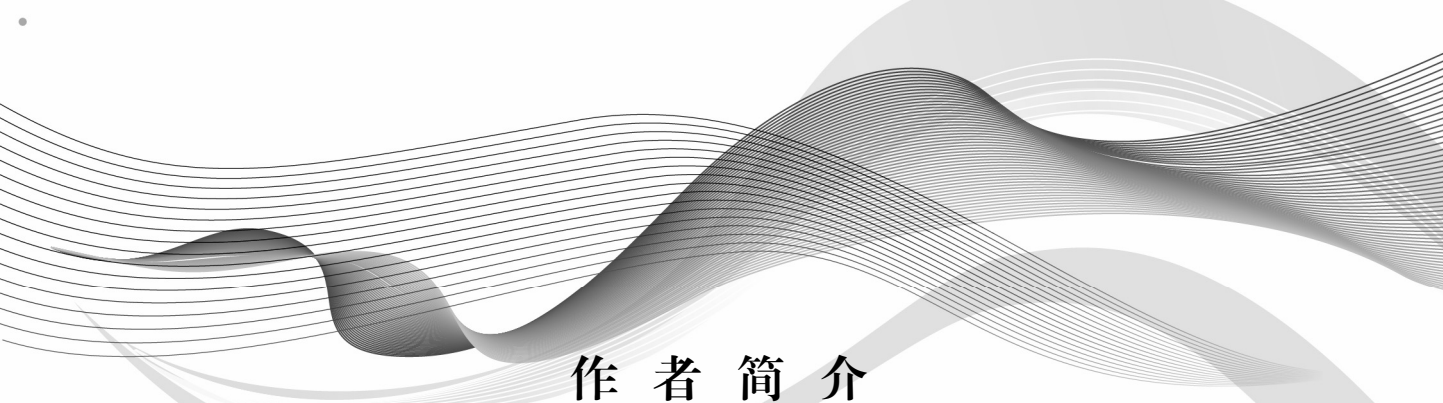
印 次: 2014 年 11 月第 1 次印刷

印 数: 3000 册 定价: 53.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888。


质量投诉请发邮件至 zlts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线: (010) 88258888。



作者简介

陆哲明，男，工学博士，浙江大学航空航天学院教授、博导、副所长。1974 年生。分别于 1995 年、1997 年和 2001 年获得哈尔滨工业大学学士、硕士和博士学位。1999 年留哈尔滨工业大学任讲师，2000 年破格为副教授，2003 年破格为教授，2004 年评为博士生导师。2004 年 10 月—2006 年 1 月作为洪堡学者赴德国弗赖堡大学作图像检索方向的访问研究。2006 年 1 月回国在哈尔滨工业大学深圳研究生院视觉信息分析与处理研究中心任主任、教授、博导；2007 年 2 月作为百人计划引进到中山大学信息科学与技术学院任教授、博导；2009 年 1 月引进到浙江大学航空航天学院航天电子工程研究所任教授、博导、副所长。陆博士为 2002 年哈尔滨市青年科技奖获得者、2003 年全国优秀博士学位论文奖获得者、2004 年教育部新世纪人才获得者、2005 年德国洪堡学者、2006 年深圳市特殊津贴专家、2011 年浙江省自然科学基金杰出青年基金获得者。陆博士长期从事多媒体信号处理、信息隐藏、复杂网络三个领域研究。这三个方面的研究工作并不是孤立的，都是在数字媒体和网络技术飞速发展的背景下展开的。截至 2014 年 1 月，陆博士在上述领域共主持省部级以上项目 12 个，共发表 SCI 检索论文 104 篇，EI 检索论文 146 篇，出版专著/教材 8 部，获省部级科技一等奖 1 项、二等奖 3 项、三等奖 1 项，厅级科技一等奖 2 项，发明专利授权 1 项。陆博士的主要学术兼职如下：国家国防科技工业局 CCSDS 工作专家组成员、IEEE 高级会员、教育部国家科学技术奖励评审专家、国家自然科学基金评审专家、SCI 国际期刊 KSII Transactions on Internet and Information Systems 编委、EI 国际期刊 Journal of Information Hiding and Multimedia Signal Processing 编委、EI 国际期刊 Information Technology Journal 编委、国际期刊 Research Journal of Information Technology 编委、国际期刊 Journal of Artificial Intelligence 编委、IEICE 会员、IIHMSP 和 IMCCC 国际会议程序委员会主席、IEEE Trans. Multimedia 和 Image Processing 等六个 IEEE 期刊审稿人、IET Image Processing 和 Electronics Letters 等四个 IET 期刊审稿人。



总 序

信息化是世界经济和社会发展的必然趋势。近年来，在党中央、国务院的高度重视和正确领导下，我国信息化建设取得了积极进展，信息技术对提升工业技术水平、创新产业形态、推动经济社会发展发挥了重要作用。信息技术已成为经济增长的“倍增器”、发展方式的“转换器”、产业升级的“助推器”。

作为国家信息化领导小组的决策咨询机构，国家信息化专家咨询委员会按照党中央、国务院领导同志的要求，就我国信息化发展中的前瞻性、全局性和战略性的问题进行调查研究，提出政策建议和咨询意见。信息化所具有的知识密集的特点，决定了人力资本将成为国家在信息时代的核心竞争力。大量培养符合中国信息化发展需要的人才 是国家信息化发展的一个紧迫需求，也是我国推动经济发展方式转变，提高在信息时代参与国际竞争比较优势的关键。2006 年 5 月，我国公布《2006—2010 年国家信息化发展战略》，提出“提高国民信息技术应用能力，造就信息化人才队伍”是国家信息化推进的重点任务之一，并要求构建以学校教育为基础的信息化人才培养体系。

为了促进上述目标的实现，国家信息化专家咨询委员会致力于通过讲座、论坛、出版等各种方式推动信息化知识的宣传、教育和培训工作。2007 年，国家信息化专家咨询委员会联合教育部、原国务院信息化工作办公室成立了“信息化与信息社会”系列丛书编委会，共同推动“信息化与信息社会”系列丛书的组织编写工作。编写该系列丛书的目的是，力图结合我国信息化发展的实际和需求，针对国家信息化人才教育和培养工作，有效梳理信息化的基本概念和知识体系，通过高校教师、信息化专家、学者与政府官员之间的相互交流和借鉴，充实我国信息化实践中的成功案例，进一步完善我国信息化教学的框架体系，提高我国信息化图书的理论和实践水平。毫无疑问，从国家信息化长远发展的角度来看，这是一项带有全局性、前瞻性和基础性的工作，是贯彻落实国家信息化发展战略的一个重要举措，对于推动国家的信息化人才教育和培养工作，加强我国信息化人才队伍的建设具有重要意义。

考虑到当时国家信息化人才培养的需求，各个专业和不同教育层次（博士生、硕士生、本科生）的需要，以及教材开发的难度和编写进度时间等问题，“信息化与信息社会”系列丛书编委会采取了集中全国优秀学者和教师，分期分批出版高质量的信息化教育丛书的方式，结合高校专业课程设置情况，在“十一五”期间，先后组织出版了“信息管理与信息系统”、“电子商务”、“信息安全”三套本科专业高等学校系列教材，受到高校相关专业学科以及相关专业师生的热烈欢迎，并得到业内专家和教师的一致好评和高度评价。

但是，随着时间的推移和信息技术的快速发展，上述专业的教育面临着持续更新、不断完善的迫切要求，日新月异的技术发展及应用变迁也不断对新时期的建设和人才培养提出新要求。为此，“信息管理与信息系统”、“电子商务”、“信息安全”三个专业教育需以综合的视角和发展的眼光不断对自身进行调整和丰富，已出版的教材内容也需及时进行

更新和调整，以满足需求。

这次，高等学校“信息管理与信息系统”、“电子商务”、“信息安全”三套系列教材的修订是在涵盖第1版主题内容的基础上进行的更新和调整。我们希望在内容构成上，既保持原第1版教材基础的经典内容，又要介绍主流的知识、方法和工具，以及最新的发展趋势，同时增加部分案例或实例，以及新的分册，使每一本教材都有明确的定位，分别体现“信息管理与信息系统”、“电子商务”、“信息安全”三个专业领域的特征，并在结合我国信息化发展实际特点的同时，选择性地吸收国际上相关教材的成熟内容。

对于这次三套系列教材（以下简称系列教材）的修订，我们仍提出了基本要求，包括信息化的基本概念一定要准确、清晰，既要符合中国国情，又要与国际接轨；教材内容既要符合本科生课程设置的要求，又要紧跟技术发展的前沿，及时地把新技术、新趋势、新成果反映在教材中；教材还必须体现理论与实践的结合，要注意选取具有中国特色的成功案例和信息技术产品中的应用实例，突出案例教学，力求生动活泼，达到帮助学生学以致用目的，等等。

为力争修订教材达到我们一贯秉承的精品要求，“信息化与信息社会”系列丛书编委会采用了多种手段和措施保证系列教材的质量。首先，在确定每本教材的第一作者的过程中引入了竞争机制，通过广泛征集、自我推荐和网上公示等形式，吸收优秀教师、企业人才和知名专家参与写作；其次，将国家信息化专家咨询委员会有关专家纳入到各个专业编委会中，通过召开研讨会和广泛征求意见等多种方式，吸纳国家信息化一线专家、工作者的意见和建议；再次，要求各专业编委会对教材大纲、内容等进行严格的审核，并对每本教材配有一至两位审稿专家。

我们衷心期望，系列教材的修订能对我国信息化相应专业领域的教育发展和教学水平的提高有所裨益，对推动我国信息化的人才培养有所贡献。同时，我们也借系列教材修订出版的机会，向所有为系列教材的组织、构思、写作、审核、编辑、出版等做出贡献的专家学者、教师和工作人员表达我们最真诚的谢意！

应该看到，组织高校教师、专家学者、政府官员以及出版部门共同合作，编写尚处于发展动态之中的新兴学科的高等学校教材，有待继续尝试和不断总结经验，也难免会出现这样那样的缺点和问题。我们衷心希望使用该系列教材的教师和学生能够不吝赐教，帮助我们不断地提高系列教材的质量。

曲作波

2013年11月1日

序 言

“十一五”期间，由国家信息化专家咨询委员会牵头，教育部信息安全专业类教学指导委员会有关领导、学者组织，众多信息安全专业著名专家和教师参与开发，并由电子工业出版社出版的“高等学校信息安全专业系列教材”，由于在体系设计上较全面地覆盖了新时期信息安全专业教育的各个知识层面，包括宏观视角上对信息化大环境下信息安全相关知识的综合介绍，对信息安全应用发展前沿的深入剖析，以及对信息安全系统建设各项核心任务的系统讲解和对一些重要信息安全应用形式的讨论，在“高等学校信息安全专业系列教材”面市后，受到高校该专业学科及相关专业师生的热烈欢迎，得到业内专家和教师的好评和高度评价，被誉为该学科专业教材中的精品系列教材。

但是，随着信息技术的快速发展，信息安全专业教育面临着持续更新、不断完善的迫切要求，其日新月异的技术发展及应用变迁也不断对新时期信息安全建设和人才培养提出新的要求。为此，信息安全专业教育需以综合的视角和发展的眼光不断对教学内容进行调整和丰富，已出版的教材内容也需及时进行更新和修改，以满足需求。

这次修订，除对“高等学校信息安全专业系列教材”第1版各册教材的主题内容进行了相应更新和调整外，同时对系列教材的总体架构进行了调整并增加了3个分册，即《信息安全数学基础》、《信息安全实验教程》和《信息隐藏概论》。

调整后的教材在体系架构和内容构成上既保持了基础的经典内容，又介绍了主流的知识、方法和工具，以及最新发展趋势，同时增加了部分案例或实例。使得系列中的每一本教材都有明确的定位，充分体现了国家“信息安全”的领域特征，在结合我国信息安全实际特点的同时，还注重借鉴国际上相关教材中适于作为信息安全本科教育知识的成熟内容。

我们希望这套修订教材能够成为新形势下高等学校信息安全专业的精品教材，成为高等学校信息安全专业学生循序渐进了解和掌握专业知识不可或缺的教科书和知识读本，成为国家信息安全新环境下从业人员及管理者学习信息安全知识的有益参考书。

高等学校信息安全专业系列教材编委会
2013年10月于北京



前言

随着计算机网络技术和多媒体处理技术的迅速发展,信息技术已经广泛应用于社会生产的各个领域。数字多媒体信息(如声音、图像、视频等)可以通过互联网快捷而有效地进行传播,这不但满足了人们的生产生活需要,也为资源共享提供了条件。然而,信息在网络传输中也存在安全隐患,例如:数字多媒体作品的版权容易受到侵犯和伪造,内容容易受到非法盗用和篡改;各种机密信息,如信用卡账号、个人隐私等容易受到非法截获和查看等。如今,信息产业已经成为国民经济中的一个重要组成部分,是社会发展的重大战略资源,信息安全已经成为影响国家安全、社会稳定、经济发展、个人利益的重大关键问题。

广义上,凡是涉及信息的安全性、完整性、可用性、真实性和可控性的相关理论和技术都是信息安全所要研究的领域。狭义上,信息安全是指信息内容的安全性,即保护信息的机密性、真实性和完整性,避免攻击者利用系统的安全漏洞进行窃听、冒充、诈骗、盗用等有损合法用户利益的行为,保护合法用户的利益和隐私。当前,密码术(Cryptography)仍然是信息安全的核心技术,基本思想是利用单钥或双钥密码算法把明文变换成密文后通过公开信道传输到接收者手中。也就是说密码术的实质是通过打乱信息内容,使其看上去像随机乱码来达到保护信息内容的目的。然而,密码术有一个致命的缺点,就是它明确地提示攻击者哪些是重要信息,因而容易引起攻击者的好奇和注意,并有被破解的可能性,而且一旦加密文件被破解后其内容就完全透明了。而且,攻击者可以在破译失败的情况下将信息破坏,使得即使是合法的接收者也无法阅读信息内容。在硬件技术高速发展以及基于互联网的强大的并行计算逐步发展的今天,传统的密码术正经受着巨大的冲击。

那么,什么技术能够弥补密码术的缺陷呢?既然乱码能激起攻击者破译热情,能不能让攻击者看不到乱码呢?答案是肯定的。人们很快想到了“暗度陈仓”的方法:明着传输无关紧要的载体而实际上暗藏着重要的秘密信息,这就引出了信息安全领域的另一个重要分支——信息隐藏(Information Hiding),其标志性里程碑是1996年第一届国际信息隐藏学术研讨会在英国剑桥的胜利举行。这样,人们开始构建以信息隐藏为核心的全新信息安全概念:信息隐藏就是将秘密信息隐藏于另一非机密的载体中。载体形式可以为任何一种数字媒体,如图像、声音、视频或一般的文档等。比较而言,密码术仅仅隐藏了信息的内容,而信息隐藏不但隐藏了信息的内容而且隐藏了信息的存在。传统的以密码术为核心技术的信息安全和隐藏式信息安全技术不是互相矛盾、互相竞争的技术,而是互补的。

一般认为,现代信息隐藏技术是由古老的隐写术(Steganography)发展而来的。该词来源于希腊语,其对应的英文是“Covered Writing”。在早期,人们常用暗号来传递秘密信息。早在公元1世纪,就已经出现用隐形墨水来传递隐秘消息。公元前400年,历史学之父Herodotus在《历史》一书中记录了几个采用不同手段进行秘密信息传递的例子:把消息藏在野兔的肚子里,假扮成猎人把消息传出去,成功地躲过了敌方哨卡的检查;剃光一位奴隶的头发并把消息刺在其头皮上,等其头发长起来后把人送往另一营地,从而实

现秘密信息的传送；将秘密消息刻在木板上，然后铺上蜡，再在蜡上写些无关紧要的话送出去。中国古代也有利用藏头诗传递秘密信息的例子以及在米粒上刻字、使用隐写墨水、道士画符遇水显影或遇火显影等例子。这些古代信息隐藏的应用实例采用了各种不同的手段，目的都是为了不引起他人的注意和怀疑。这也就引出了信息隐藏的根本目的，即隐藏信息的存在。古代隐写术发展一直比较缓慢，没有成为一门独立的学科，使得人们对于信息保密更多的是采用密码术。直到信息技术和计算机技术高度发达的今天，数字化信息与隐写术相结合为古老的隐写术注入了新的活力，使得数字化信息隐藏技术成为一门全新的技术，为探索非密码的通信安全提供了新途径。现代信息隐藏技术是在 20 世纪 90 年代后才逐步发展起来的，大家普遍公认的时间起点是 1994 年，那一年 Schyndel 等人在 IEEE 国际图像处理会议上首次明确提出了“数字水印”概念。

1996 年 5 月在英国剑桥召开的第一次国际信息隐藏学术研讨会上，使得一些研究团体在信息隐藏的基本概念和术语上达成共识，从而使信息隐藏作为一门新学科开始得到快速发展。信息隐藏也称数据隐藏，它是集多学科理论与技术于一身的新兴技术，它利用人类感官对数字信号的感觉冗余，将秘密信息隐藏于另一非保密载体（如图像、视频、音频、信道甚至整个系统）中，以不引起检查者的注意。信息隐藏技术有多种含义：一是信息不可见，二是信息的存在性隐蔽，三是信息的接收方和发送方隐蔽，四是传输的信道隐蔽。信息隐藏技术包括隐蔽信道、匿名通信、隐写术、数字水印技术、数字指纹技术、闕下信道和低截获概率通信等多个学科分支。

基于信息隐藏这门新学科的发展现状，本书旨在介绍近二十年来信息隐藏领域的公认的广泛提及和深入研究的一些基础理论和典型方法，尽量以浅显易懂的方式为学习信息隐藏技术的来自信息安全专业的本科生和研究生提供入门教材。在学习本书之前，读者们需要具备微积分、概率论、信息论、正交变换和计算机科学的基础知识，了解图像处理、音频信号处理、视频信号处理、模式识别等相关概念，掌握必要的计算机编程语言和具备较好的仿真验证能力。本书一共 9 章，分别是绪论、隐写术、数字水印技术、数字指纹技术、无损信息隐藏技术、其他信息隐藏研究分支简介、隐写分析技术、数字水印攻击技术和信息隐藏技术的应用。为了让读者更好地理解信息隐藏技术，各章都配有习题。若要粗略学习信息隐藏的主要分支，建议学习第 1 章到第 4 章，建议教师各分配 2 学时、8 学时、8 学时、6 学时课堂授课，且为第 2 章和第 3 章各配 2 学时上机实验，这样一共 24 学时课堂 4 学时上机。若要对所有信息隐藏研究分支作了解，建议学习第 1 章到第 6 章，其中第 5 章分配 6 学时课堂讲授和 2 学时上机实验，第 6 章分配 2 学时课堂讲授，这样一共 32 学时课堂 6 学时上机。若要详细了解隐写术和数字水印技术的攻防两个方面及其应用，建议学习第 1、2、3、7、8、9 章这六章，其中 7、8、9 三章分别分配 6、6、2 学时课堂授课，这样一共是 32 学时课堂 4 学时上机。下面介绍各章的背景和主要内容。

信息隐藏技术是在传统密码术原理的基础上发展起来的一门涉及信息论、密码学、应用数学、计算机科学、网络技术、通信技术等多种学科的综合性学科。相对于密码术，信息隐藏技术的优点在于它隐藏了秘密信息的存在性，减小了受攻击的风险。正是这种特性使得信息隐藏技术在保护信息安全方面比密码技术具有更好的发展前景。本书的第 1 章从网络信息安全问题入手，对信息隐藏技术的概念、模型、分类、研究分支、历史发展和应用领域进行概述。

与其他信息隐藏研究分支相比，隐写术是一门古老的技术，主要用于保密通信。它将秘密信息嵌入到看上去普通的信息中进行传送，以防止第三方检测出秘密信息。隐写术

关注的重点是如何让秘密信息的存在不被发现。本书的第 2 章首先概述保密通信的有关背景，接着概述隐写术的有关概念、分类和性能评价问题，然后按照载体类型的不同分别介绍基于文本、图像、音频和视频等载体的隐写术。

信息技术和计算机网络的迅速发展，使得在网上传播的多媒体作品的版权保护和内容认证面临着日益严峻的挑战。为了保护数字媒体的知识产权，以前人们采用了将数据加密的方法，使得只有掌握密钥的授权用户才能解密数据，从而使用数字媒体产品。但这种方法只能控制用户是否能够存取数据，与数据本身并无直接关系，因此一旦被破解，这些数据就会很轻易地被修改、复制、传播。为消除这个隐患，人们又提出了新的知识产权保护手段——数字水印技术。数字水印技术是将一段标志版权所有者的信息嵌入到要保护的媒体中，但在这个过程中通常采用特定的技术手段使被嵌入的信息不会被人感知到，只有知识产权的所有者才能通过检测器确定数字水印是否存在。本书第 3 章从数字水印技术的提出背景入手，首先介绍数字水印技术的相关概念、分类、框架模型和性能评价，然后按照载体类型的不同分别介绍用于图像、音频和视频等载体的数字水印技术。

信息技术的迅猛发展及以其为基础的电子商务的广泛应用，使各类文字、图片、影视等作品通过网络的传播范围空前扩大，为创作者和发行商带来了新机遇。但同时，人们也很容易对以数字形式存在的产品进行非法复制和分发。数字指纹技术是近几年发展起来的一种新型数字版权保护技术，它的原理是版权者在其分发的数字作品复制中嵌入与用户身份相关的唯一信息，当发现非法复制时，版权者凭借嵌入信息可以识别非法分发复制的用户，进而通过法律诉讼和惩罚来达到保护版权权益、对非法行为进行威慑的目的，可以看出数字指纹技术实现了一种版权跟踪机制。本书第 4 章从数字指纹技术的提出背景入手，首先介绍数字指纹技术的相关概念、分类、框架模型和性能评价，然后概述最重要的指纹编码和指纹协议问题，接着详细介绍基于不同指纹协议的数字指纹技术，最后讨论抗共谋攻击的指纹编码问题。

传统信息隐藏技术通常给载体对象引入了一些细微的、不可逆的或一些永久的失真，尽管这些失真是非常轻微的，但是在一些对精度要求高的领域，比如法律、医学和军事系统等，当秘密信息被提取后，需要无失真地恢复原始载体，因此即使是非常轻微的失真也是不允许的。在这种情况下，出现了无失真恢复原始载体的嵌入技术，称为无损信息隐藏技术。本书的第 5 章从无损信息隐藏技术的提出背景入手，首先介绍无损信息隐藏技术的相关概念、分类、框架模型和性能评价，接着以图像为载体对象介绍基于不同嵌入域（空域、整数变换域和压缩域）的无损信息隐藏技术。

除了隐写术、数字水印技术和数字指纹技术外，信息隐藏领域还有一些其他研究分支，例如：普遍存在于安全操作系统、安全网络、安全数据库系统中的危害系统安全策略的隐蔽信道；针对公钥密码技术的数字签名、认证等应用密码体制的闕下信道等。为了让读者对信息隐藏的各个分支都有所了解，本书第 6 章分别简要介绍隐蔽信道、闕下信道、低截获概率通信和匿名通信等研究分支。

隐写术为大众在机密性和个人隐私性方面提供了保护，同时也使得恐怖组织或谍报机构进行非法信息传递变得更为便利。如果不当地使用隐写术，会损害企业 and 国家利益，给公共安全和社会稳定带来威胁。在检测隐藏信息、监控和阻截非法隐蔽通信方面，各国军方和安全部门表现出了十分迫切的需求，隐写分析术（Steganalysis）应运而生。隐写分析术是对数字媒体信号进行统计分析，判断其中是否藏有秘密信息的技术。更高层次的隐写分析术可以对秘密信息的长度、隐藏位置等进行判断。隐写分析的目的是检测、阻截和

破坏隐蔽通信。隐写分析术和隐写术是两种相互对抗且相互促进的技术。本书第 7 章首先介绍了隐写分析的基本概念、分类和评价指标，然后分别介绍针对图像、音频和视频载体的隐写分析技术，其中以图像载体为主。针对每种载体，主要分专用分析法和通用分析法分别进行介绍。

数字水印作为一种数据认证和版权保护的手段，必然会受到各种形式的攻击。研究数字水印攻击的目的是发现现有技术的漏洞和缺陷，以提出对策来提升未来水印设计的抗攻击能力。本书第 8 章首先给出攻击的定义、分类和相关概念，然后介绍针对安全性的三种攻击技术，最后介绍针对应用系统而与水印算法安全性及鲁棒性无关的系统攻击技术。

信息隐藏技术已经在许多领域得到了广泛应用，主要集中在如下几大方面：① 数据保密通信；② 身份认证；③ 数字作品的版权保护与盗版追踪；④ 完整性、真实性鉴定与内容恢复。本书第 9 章首先介绍信息隐藏技术的四个最重要的应用领域：知识产权保护、军事保密通信、交易跟踪和真伪鉴别。然后介绍复制控制、广播监控、设备控制及其他一些应用领域。

本书可作为高等院校具有一定计算机基础的信息安全专业、电子信息工程专业、计算机专业、通信工程专业的研究生或高年级本科生的教材或参考书，也可作为科研院所相关专业的科技工作者的参考书。本书的第 1、2、3、5、7、8 章由陆哲明教授执笔，第 4、6 章由聂廷远副教授执笔，第 9 章由吉爱国教授执笔，最后由陆哲明教授审定。本书广泛参考了国内外信息隐藏研究领域的学术论文、学位论文和学术著作，并包含了作者的部分研究成果，这些成果得到了多个国家自然科学基金项目（No. 61171150、No. 61003255 和 No. 60272074）和浙江省杰出青年基金项目（No. R1110006）的资助，在此致以深深的谢意。在本书的撰写过程中还得到了浙江大学航空航天学院航天电子工程研究所、青岛理工大学信息对抗研究所各位教师、博士生和硕士生的协助，在此表示衷心的感谢。

限于水平，书中难免有错误与不妥之处，恳请读者批评指正。

陆哲明

于杭州浙江大学航空航天学院航天电子工程研究所

聂廷远，吉爱国

于青岛理工大学信息对抗研究所

2014 年 2 月

目 录

第 1 章 绪论	1
1.1 网络信息安全	2
1.1.1 网络时代和信息安全问题	2
1.1.2 信息安全技术概述	2
1.2 信息隐藏的基本概念	4
1.2.1 信息隐藏的产生背景	4
1.2.2 信息隐藏的定义和相关术语	4
1.2.3 信息隐藏技术的特性和要求	6
1.3 信息隐藏的模型	7
1.3.1 囚徒模型	7
1.3.2 通用模型	8
1.3.3 通信模型	9
1.3.4 广义模型	9
1.3.5 不对称信息空间模型	10
1.4 信息隐藏的研究分支	12
1.4.1 隐写术	12
1.4.2 版权标记	13
1.4.3 隐蔽信道	13
1.4.4 阙下信道	14
1.4.5 低截获概率通信	14
1.4.6 匿名通信	15
1.5 信息隐藏技术的分类	15
1.5.1 按载体类型分类	15
1.5.2 按密钥对称性分类	17
1.5.3 按嵌入域分类	18
1.5.4 其他分类方式	19
1.6 信息隐藏技术的历史发展	19
1.6.1 古代信息隐藏技术	20
1.6.2 近代信息隐藏技术	22
1.6.3 现代数字信息隐藏技术	23
1.7 信息隐藏的应用领域	25
1.7.1 保密通信	25
1.7.2 版权保护和复制控制	25
1.7.3 数字指纹（盗版者/叛逆者追踪）	26

1.7.4	内容认证（真伪鉴别、完整性鉴别）	26
1.7.5	标注	27
1.7.6	其他应用	27
1.8	本章小结	28
	习题	28
第 2 章	隐写术	31
2.1	保密通信概述	32
2.1.1	基本概念和分类	32
2.1.2	基于经典密码术的保密通信	32
2.1.3	混沌保密通信	33
2.1.4	量子保密通信	34
2.1.5	基于隐写术的保密通信	34
2.2	隐写术的相关概念和分类	34
2.2.1	隐写术的基本概念	34
2.2.2	隐写术的分类	35
2.2.3	语义隐写术概述	36
2.2.4	技术隐写术概述	38
2.3	隐写系统的性能评价	40
2.3.1	透明性	40
2.3.2	秘密信息的正确恢复率（鲁棒性）	42
2.3.3	隐写容量	42
2.3.4	安全性	43
2.3.5	系统复杂度	43
2.4	基于文本载体的隐写术	44
2.4.1	引言	44
2.4.2	基于文档格式微调的隐写术	44
2.4.3	基于空格和标点符号的隐写术	45
2.4.4	基于字符特征的隐写术	46
2.4.5	基于自然语言的隐写术	48
2.4.6	基于变换域的隐写术	50
2.4.7	对比和总结	50
2.5	基于图像载体的隐写术	50
2.5.1	引言	50
2.5.2	空域隐写术	53
2.5.3	变换域隐写术	60
2.5.4	JPEG 图像隐写术	67
2.6	基于音频载体的隐写术	74
2.6.1	引言	74
2.6.2	时域隐写方法	76
2.6.3	变换域隐写方法	79

2.6.4	压缩域隐写方法	81
2.7	基于视频载体的隐写术	82
2.7.1	引言	82
2.7.2	未压缩视频中的隐写	84
2.7.3	压缩视频中的隐写	85
2.7.4	分析与比较	85
2.8	本章小结	86
	习题	87
第 3 章	数字水印技术	89
3.1	数字水印技术的提出背景	90
3.2	数字水印技术的相关概念和分类	91
3.2.1	数字水印技术相关概念	91
3.2.2	数字水印技术和隐写术的区别	92
3.2.3	数字水印及数字水印技术的分类	93
3.3	数字水印系统的框架模型	94
3.3.1	数字水印系统基本框架	94
3.3.2	基于通信系统的数字水印模型	95
3.3.3	数字水印系统的几何模型	97
3.4	数字水印技术的应用和性能评价	101
3.4.1	数字水印技术的应用	102
3.4.2	数字水印技术的特性	103
3.4.3	数字水印系统的评价问题	105
3.5	数字图像水印技术	110
3.5.1	数字图像水印系统的基本要求	110
3.5.2	数字图像水印系统的基本模型和算法分类	111
3.5.3	数字图像水印系统的关键技术	112
3.5.4	数字图像水印算法的评价	118
3.5.5	典型鲁棒图像水印算法	122
3.5.6	典型脆弱图像水印算法	125
3.6	数字音频水印技术	126
3.6.1	音频水印系统的基本要求	126
3.6.2	音频水印系统的基本模型	127
3.6.3	含水印音频质量的评价	128
3.6.4	数字音频水印算法的鲁棒性评测	130
3.6.5	典型时域数字音频水印算法	134
3.6.6	典型变换域数字音频水印算法	134
3.6.7	典型压缩域数字音频水印算法	137
3.7	数字视频水印技术	139
3.7.1	数字视频水印技术的特点和面临的挑战	139
3.7.2	视频数字水印系统的模型和算法分类	141

3.7.3	典型原始域视频水印算法	143
3.7.4	典型压缩域视频水印算法	150
3.8	本章小结	152
	习题	153
第 4 章	数字指纹技术	155
4.1	数字指纹技术的提出背景	156
4.1.1	指纹和指纹识别	156
4.1.2	数字指纹技术的提出背景	156
4.2	数字指纹技术的相关概念和分类	157
4.2.1	数字指纹技术的相关概念	158
4.2.2	数字指纹技术的特性要求	159
4.2.3	数字指纹技术的分类	161
4.3	数字指纹系统模型和性能评价	162
4.3.1	数字指纹系统模型	162
4.3.2	数字指纹系统的攻击手段	163
4.3.3	数字指纹技术的性能评价	165
4.4	指纹编码和指纹协议概述	168
4.4.1	指纹编码概述	168
4.4.2	指纹协议概述	171
4.5	统计指纹技术	174
4.6	对称指纹技术	175
4.6.1	对称指纹技术的基本方案	175
4.6.2	基本叛逆者追踪协议	176
4.7	非对称指纹技术	179
4.7.1	非对称指纹技术的基本方案	180
4.7.2	非对称叛逆者追踪协议	181
4.8	匿名指纹技术	182
4.8.1	匿名指纹技术的基本思想	182
4.8.2	一种典型的匿名指纹技术	183
4.9	共谋安全指纹技术	186
4.9.1	共谋攻击方式	186
4.9.2	术语与定义	188
4.9.3	编码设计	189
4.9.4	©-安全码	190
4.9.5	I 码	191
4.9.6	BIBD 码	191
4.9.7	典型叛逆者追踪方案	193
4.10	本章小结	195
	习题	195

第 5 章 无损信息隐藏技术	197
5.1 无损信息隐藏技术的提出背景	198
5.2 无损信息隐藏技术的相关概念和分类	198
5.2.1 无损信息隐藏的相关概念	198
5.2.2 无损信息隐藏的关键问题	199
5.2.3 无损信息隐藏的分类	200
5.3 无损信息隐藏系统的框架和性能评价	201
5.3.1 无损信息隐藏的框架模型	201
5.3.2 无损信息隐藏算法的评价	201
5.4 空域无损信息隐藏技术	203
5.4.1 基于无损压缩替换的无损信息隐藏	204
5.4.2 基于模加的无损信息隐藏	206
5.4.3 基于差值扩展的无损信息隐藏	207
5.4.4 基于直方图移位的无损信息隐藏	210
5.5 变换域无损信息隐藏技术	211
5.5.1 整数 DCT 变换域数值扩展技术	211
5.5.2 整数 DWT 变换域数值扩展技术	214
5.5.3 整数变换域直方图移位技术	217
5.6 压缩域无损信息隐藏技术	219
5.6.1 压缩域无损信息隐藏技术和应用要求	219
5.6.2 BTC 压缩域无损信息隐藏	220
5.6.3 JPEG 压缩域无损信息隐藏	225
5.7 本章小结	234
习题	234
第 6 章 其他信息隐藏研究分支简介	237
6.1 隐蔽信道	238
6.1.1 隐蔽信道基本概念	238
6.1.2 隐蔽信道分类	239
6.1.3 隐蔽信道研究领域	240
6.1.4 隐蔽信道分析技术	242
6.2 阈下信道	243
6.2.1 阈下信道相关概念	243
6.2.2 阈下信道的存在性	245
6.2.3 阈下信道的模型和评价指标	246
6.2.4 阈下信道的构造方法	248
6.3 低截获概率通信	251
6.3.1 扩频通信技术	251
6.3.2 流星余迹猝发通信技术	256
6.4 匿名通信	258
6.4.1 匿名通信概述	258

6.4.2	匿名通信系统体系结构	259
6.4.3	匿名性能与效率	262
6.5	本章小结	264
	习题	264
第 7 章	隐写分析技术	265
7.1	隐写分析基本概念和分类	266
7.1.1	基本概念	266
7.1.2	分类	266
7.2	隐写分析算法的评价指标	269
7.2.1	可靠性和准确性	269
7.2.2	适用性	269
7.2.3	分类代价	271
7.2.4	实用性和计算复杂度	272
7.3	图像专用隐写分析	273
7.3.1	引言	273
7.3.2	针对空域 LSB 替换的专用隐写分析算法	274
7.3.3	针对 LSB 匹配隐写的专用隐写分析算法	279
7.3.4	针对 JPEG 图像隐写的专用分析算法	281
7.4	图像通用隐写分析	285
7.4.1	图像通用隐写分析基本思想	285
7.4.2	支持向量机分类技术	286
7.4.3	图像通用隐写分析算法概述	290
7.4.4	典型通用隐写分析算法	293
7.5	音频隐写分析技术	297
7.5.1	专用音频隐写分析的机理	297
7.5.2	通用音频隐写分析的机理	299
7.5.3	音频隐写分析系统模型	303
7.5.4	针对 LSB 隐藏的隐写分析方法	304
7.5.5	通用音频隐写分析方法	308
7.6	视频隐写分析技术	309
7.6.1	视频隐写分析基本原理及框架	309
7.6.2	视频隐写分析特点	310
7.6.3	视频隐写分析算法的评估指标	311
7.6.4	强针对性视频隐写分析算法	312
7.6.5	视频盲隐写分析算法	315
7.7	本章小结	316
	习题	316
第 8 章	数字水印攻击技术	319
8.1	数字水印系统的鲁棒性和安全性	320
8.1.1	鲁棒性	320

8.1.2	安全性	320
8.2	数字水印攻击技术的相关概念和分类	322
8.2.1	攻击方法的分类	322
8.2.2	受限的水印操作	324
8.2.3	关于对手的假设	325
8.3	非授权去除攻击	326
8.3.1	引言	326
8.3.2	去除攻击	327
8.3.3	掩盖攻击	331
8.3.4	对策	334
8.4	非授权嵌入攻击	336
8.4.1	引言	336
8.4.2	复制攻击	336
8.4.3	多重嵌入攻击	337
8.4.4	协议攻击	337
8.4.5	针对脆弱水印的非授权嵌入攻击	339
8.4.6	对策	340
8.5	非授权检测攻击	343
8.5.1	问题	343
8.5.2	对策	343
8.6	系统攻击和法律攻击	344
8.6.1	引言	344
8.6.2	体系结构问题	345
8.6.3	典型合法攻击	346
8.7	本章小结	347
	习题	347
第 9 章	信息隐藏技术的应用	349
9.1	知识产权保护	350
9.1.1	基于数字水印的版权保护系统框架	350
9.1.2	数字水印版权系统与 PKI / CA 体系的对比	353
9.1.3	系统实现方案	353
9.2	军事保密通信	355
9.2.1	系统组成	355
9.2.2	系统采用的秘密语音隐藏方案	356
9.2.3	隐藏效果	357
9.3	交易跟踪	358
9.3.1	系统整体架构	358
9.3.2	版权注册与认定	360
9.3.3	指纹证书管理	361
9.3.4	指纹嵌入与作品分发	362

9.3.5	作品版权主动追踪·····	362
9.4	真伪鉴别·····	363
9.4.1	电子公文特点·····	363
9.4.2	嵌有水印公章的特点和需求·····	364
9.4.3	公章认证方案·····	365
9.4.4	电子印章系统·····	365
9.5	复制控制·····	368
9.5.1	现有的复制控制技术·····	368
9.5.2	视频水印技术用于复制控制需解决的问题·····	370
9.5.3	一种典型的 DVD 复制控制系统方案·····	370
9.6	广播监控·····	374
9.6.1	数字视频广播·····	374
9.6.2	视频广播监控问题·····	375
9.6.3	基于数字水印的视频广播监控·····	377
9.7	其他应用·····	378
9.7.1	印刷防伪·····	378
9.7.2	软件保护·····	379
9.7.3	媒体桥·····	380
9.8	本章小结·····	380
	习题·····	380
	参考文献·····	381
	索引·····	389

绪 论

本章引言

计算机和网络技术的迅速发展,促使信息与网络安全成为新时代下的关键问题。作为信息安全的一个重要分支,信息隐藏技术在近年来得到迅速发展,尤其是在数字版权保护、签名认证、篡改检测、保密通信等领域都有信息隐藏的具体应用。随着信息隐藏技术的发展和研究的深入,对相关基础理论和框架的研究需求也逐步增加。本章从网络信息安全问题入手,对信息隐藏技术的概念、模型、分类、研究分支、历史发展和应用领域进行概述。

本章重点

- 信息隐藏的基本概念;
- 信息隐藏的模型;
- 信息隐藏的研究分支;
- 信息隐藏的分类。



1.1 网络信息安全

1.1.1 网络时代和信息安全问题

人类文明的发展进程与网络息息相关。早在远古时期，人类就构造出山间小路和林中小径并连成网络；在农业社会，人类又构造出用于灌溉的各种水利网络，通过航海网络，资本主义才走遍全世界；在工业社会，航空网、公路网、铁路网、电话网和商业网极大方便了人们的交流和贸易。从 20 世纪 90 年代开始，随着计算机技术和通信技术的迅猛发展，出现了**国际互联网（Internet）**，人类真正步入了网络时代。**网络时代（Networking Era）**是指在电子计算机和现代通信技术相互结合基础上构建的宽带、高速、综合、广域型数字化电信网络的时代。网络时代是人类文明的一个重要里程碑。历史上，从没有其他网络像 Internet 那样，在如此短的时间内影响如此多的政府、企业和个人。以中国网民数量为例，1997 年底才 62 万，到 2002 年初达 1591 多万，而到了 2013 年 6 月则已达 5.91 亿。Internet 已经打破了传统的边界概念，并渗透到了政治、军事、金融、商业、交通、电信、文教等各个领域，全球经济一体化和信息网络化相互促进、相互依存的趋势越来越明显。例如，电子商务作为网络时代经济活动全新的技术手段和方法，已成为 Internet 最广阔的应用领域，各国政府也纷纷构建电子政府以适应网络时代要求。

信息安全（Information Security）问题随着互联网的发展伊始便逐步显现出来，而随着网络技术和信息技术的不断发展，该问题日显突出。当前互联网信息安全问题已经渗透到各个领域，隐私泄漏、企业数据外泄、黑客攻击以及商业间谍等已经让人们不堪其扰。在全球化快速发展的今天，互联网的触角延伸到世界各国的各个角度，在网络安全跨越国界、网络空间战不断升级的情况下，信息安全也自然上升到国家安全的高度。由于网络安全问题不仅关系到公民个人财产和人身安全，也关系到国家安全和社会稳定，如何确保信息系统的安全已成为全社会关注的问题。根据中国互联网络信息中心发布的《2012 年中国网民信息安全状况研究报告》显示，多年来，我国不断加强网民的信息安全治理，但网络信息安全形势仍然极为严峻。主要问题如下：① 新型的信息安全事件不断出现，且迅速向更多网民蔓延；② 导致信息安全事件的情境日益多样复杂化，令网民防不胜防；③ 信息安全所引起的直接经济损失已达到较大规模，接近 200 亿元；④ 发起信息安全事件的因素已从以前的好奇心理升级为明显的逐利性，经济利益链条已经形成；⑤ 信息安全事件中所涉及的信息类型、危害类型越来越多，且涉及网民的隐私，潜在后果更严重。

网络信息时代的到来也对**版权保护（Copyright Protection）**提出了新的挑战。所谓**版权（Copyright）**，有时也称作者权，在我国被称为著作权，是基于特定作品的精神权利以及全面支配该作品并享受其利益的经济权利的合称。随着网络规模的不断扩大和数字化技术的不断成熟，网上各种数字化图书、报刊杂志、绘画、照片、音乐、歌曲及影视作品的数量也急剧增加。这些数字化产品和服务都可实现网络传送，不受时间和空间限制，甚至无须物流传输，在交易和支付完成后，就可高效快捷地通过网络提供给客户。而网络的开放性和资源共享使得如何有效地保护网络数字化产品的版权成为一个十分重要的问题。必须采取行之有效的技术手段以防止数字化产品的篡改、假冒、剽窃和盗用等。

1.1.2 信息安全技术概述

信息安全技术（Information Security Technology）是一门综合的学科，它涉及**信息论**

(Information Theory)、**计算机科学** (Computer Science) 和**密码术** (Cryptography) 等多方面知识, 主要研究计算机系统和通信网络内信息的保护方法以确保系统内信息的安全、保密、真实和完整。网络信息安全涉及信息传输安全、信息存储安全、对网络传输信息内容的审计以及对用户的鉴别和授权四方面。为保障数据传输的安全, 需用数据加密技术、**完整性** (Integrity) 鉴别技术; 为保证信息存储的**安全性** (Security), 需保障数据库安全和终端安全; 信息内容审计, 是对进出内部网络的信息进行实时内容审计, 以防止或追查可能的泄密行为。对用户的鉴别是对网络中的主体进行验证的过程, 通常采用三种方法验证主体身份: 一是只有该主体了解的秘密, 如口令、密钥; 二是主体携带的物品, 如智能卡和令牌卡; 三是只有该主体具有的独一无二的特征或能力, 如指纹、声音、视网膜或签字等。

网络信息安全的技术特征主要表现在以下几个方面。① **完整性**: 是网络信息未经授权不能改变的特性。即对抗主动攻击, 保证数据的一致性, 防止数据被非法用户修改和破坏。② **保密性** (Confidentiality): 是网络信息不泄露给未经授权者的特性。即对抗被动攻击, 以保证**秘密信息/消息** (Secret Information/Message) 不会泄露给非法用户。③ **可用性** (Availability): 是网络信息可被授权者访问并按需使用的特性。即保护合法用户对信息和资源的使用不会被不合理拒绝。④ **不可否认性** (Non-repudiation): 也称为不可抵赖性, 即网络上所有参与者都不能否认或抵赖曾经完成的操作和承诺。发送方不能否认已发送的信息, 接收方也不能否认已收到的信息。⑤ **可控性** (Controllability): 是对网络信息的传播及内容具有控制能力的特性。即能够对网络信息实施安全监控。

保护信息安全所采用的手段称作**安全机制** (Security Mechanism)。所有安全机制都是针对某些安全攻击威胁而设计的, 可以按照不同的方式单独或组合使用。网络中采用的安全机制主要有以下几种。① **信息加密和隐藏机制**: 加密使攻击者无法读懂消息的内容从而保护信息; 而隐藏则是将有用的信息隐藏在其他信息中, 使攻击者无法发现, 不仅实现了信息的保密, 也保护了通信本身。至今, 信息加密仍是保障信息安全的最基本的手段。**信息隐藏** (Information Hiding) 则是信息安全领域的一个新方向, 它在数字产品版权保护等领域的应用中正越来越受到人们的重视。② **完整性保护**: 用于防止非法篡改, 利用密码术的完整性保护能够很好地对付非法篡改。完整性的另一用途是提供不可抵赖服务, 当信息源的完整性可以被验证却无法模仿时, 收到信息的一方可以认定信息的发送者, **数字签名** (Digital Signature) 就可以提供这种手段。③ **认证机制**: 网络安全的基本机制, 即网络设备之间应互相认证对方身份, 以保证合法的用户进行正确的操作并进行正确的审计。④ **审计**: 防止内部犯罪和事故后调查取证的基础, 通过对一些重要的事件进行记录, 从而在系统出现错误或受到攻击时能定位错误和找到攻击成功的原因。审计信息应具有防止非法删除和修改的措施。⑤ **权力控制和存取控制**: 主机系统必备的安全手段, 即系统根据正确的认证, 赋予某用户适当的操作权力, 使其不能进行越权的操作。该机制一般采用角色管理办法, 针对系统需要定义各种角色, 如经理、会计等, 然后对他们赋予不同的执行权力。⑥ **业务填充**: 在业务空闲时发送无用的随机数据, 增加攻击者通过通信流量获得信息的困难。同时也增加了加密通信的破译难度。发送的随机数据应具有良好的模拟性能, 能够以假乱真。我们可以从不同角度对信息安全作出不同的解释。**狭义信息安全** (Narrow Information Security) 是指信息的机密性、完整性和不可否认性, 主要研究加密和认证等算法。狭义信息安全还可能包括与意识形态相关的内容安全。**广义信息安全** (General Information Security) 通常是指信息在采集、

加工、传递、存储和应用等过程中的完整性、机密性、可用性、可控性和不可否认性以及意识形态相关的内容安全。一般意义上,信息安全包括内容安全和控制安全两部分。国际标准化组织定义信息安全为“为数据处理系统建立和采取的技术和管理的安全保护,保护计算机硬件、软件和数据不因偶然和恶意的原因而遭到破坏、更改和泄漏”。通常,信息安全主要从两个方面来实现。① 计算机安全:保护计算机硬件、软件和数据不因偶然或恶意的原因而遭破坏、更改和泄漏。② 网络安全:保障网络服务的可用性和网络信息的完整性。

1.2 信息隐藏的基本概念

1.2.1 信息隐藏的产生背景

数字化设备和计算机技术为多媒体信息(音频、图像、视频、动画、文本和三维模型等)的制作和存取提供了极大的便利。同时,随着因特网的日益普及,多媒体信息的交流已达到了前所未有的深度和广度。人们如今也可以通过因特网发布自己的作品、传递重要信息或进行网络贸易等,但是随之出现的问题也十分严重:作品侵权和隐私侵犯更加容易、篡改和伪造更加方便、恶意攻击更加猖獗。因此,如何有效地保护知识产权和保证网络信息安全,已受到人们的高度重视。另外,现代战争是多兵种协同作战的立体战争,在某种程度上来说是电子战、信息战和网络战,而且已从传统电子对抗转化为信息对抗。在这些背景下,一门新兴交叉学科——**信息隐藏**^[1]正式诞生了。如今信息隐藏作为**保密通信**(Secret Communication)、**知识产权保护**(Intellectual Property Right Protection)和**内容认证**(Content Authentication)等领域的主要手段,正得到广泛地研究与应用。

为了打击盗版犯罪和保证网络信息安全,一方面要通过立法来加强对知识产权和网络信息的保护,另一方面必须要用先进的技术手段来保障法律的实施。虽然**密码术**^[2]可用来解决其中的部分问题,但密码术存在如下三大缺点。① 它明确地提示攻击者哪些是重要信息,容易引起攻击者的好奇和注意,并有被破解的可能性。② 一旦加密文件被破解,其内容就完全透明了。③ 攻击者可以在破译失败的情况下将信息破坏,使得即使是合法的接收者也无法阅读信息内容。20 世纪 90 年代早期,信息隐藏技术以其特有的优势解决了密码术的一些缺陷,开始引起人们的普遍关注,而且信息隐藏的各种应用引起不同研究团体的关注和重视。1996 年 5 月第一届国际信息隐藏学术研讨会在英国剑桥的召开,使这些独立的研究团体走到一起,从而在信息隐藏的一些基本概念和术语上达到共识^[1,3,4]。下面介绍信息隐藏的定义和相关术语。

1.2.2 信息隐藏的定义和相关术语

信息隐藏也称**数据隐藏**(Data Hiding)。从广义上看,信息隐藏有多种含义,一是信息不可见,二是信息的存在性隐蔽,三是信息的接收方和发送方隐蔽,四是传输的信道隐蔽。**信息隐藏**就是将秘密信息隐藏于另一公开载体中,以不引起检查者的注意。这里的载体可以是图像、视频、音频,也可以是信道,甚至是某套编码体制或整个系统。由此可见,密码术仅仅隐藏了信息的内容,而信息隐藏技术不但隐藏了信息的内容,而且隐藏了信息的存在。广义上的信息隐藏技术包括:**隐写术**(Steganography)、**数字水印技术**(Digital Watermarking)^[5]、**数字指纹技术**(Digital Fingerprinting)、**隐蔽信道**(Covert Channel)、**阈下信道**(Subliminal Channel)^[6]、**低截获概率通信**(Low Probability Intercept

Communication) 和**匿名通信** (Anonymous Communication) 等。从狭义上看, 信息隐藏就是将某一秘密信息隐藏于另一公开的信息中, 然后通过公开信息的传输来传递秘密信息。狭义上的信息隐藏技术通常指隐写术与数字水印技术 (以及数字指纹技术)。

信息之所以能够隐藏在多媒体数据中是因为以下两点。① 多媒体信息本身存在很大的冗余性。从信息论的角度看, 未压缩的多媒体信息的编码效率是很低的, 所以将这些秘密信息嵌入到多媒体信息中进行保密传输是完全可行的, 并不会影响多媒体信息本身的传送和使用。② 人眼或人耳本身具有掩蔽效应, 比如人眼对灰度的分辨率只有几十个灰度级; 对边缘附近的信息不敏感。利用人的这些特点, 可以很好地将信息隐藏而不被察觉。

通常, 信息隐藏与信息加密都需要把对信息的保护转化为对密钥的保护。信息隐藏不同于传统的密码术。密码术主要是研究如何将秘密信息进行特殊的编码, 以形成不可识别的**密文** (Cipher Text) 进行传递; 而信息隐藏则主要研究如何将某一秘密信息隐藏于另一公开的载体中, 然后通过公开载体的传输来传递秘密信息。对加密通信而言, 可能的监测者或非法拦截者可通过截取密文, 并对其进行破译, 或将密文进行破坏后再发送, 从而影响秘密信息的安全; 但对信息隐藏而言, 可能的监测者或非法拦截者则难以从公开载体中判断秘密信息是否存在, 难以截获秘密信息, 从而能保证秘密信息的安全。为了增加破译的难度, 也可以把密码术和信息隐藏技术相结合, 即先对待嵌入对象进行加密得到密文, 再把密文隐藏到**载体对象** (Cover Object) 中。由此可见, 传统的以密码术为核心技术的信息安全和隐藏式信息安全技术并不是互相矛盾、互相竞争的技术, 而是互补的。

在广义信息隐藏概念下, 下面介绍第一届信息隐藏国际会议中给出的一些术语。人们把希望被秘密隐藏的对象称为**嵌入对象** (Embedded Object), 它是含有特定用途的**秘密信息** (秘密消息)。用于隐藏嵌入对象的非保密载体称为**载体对象**。需要指出, 对象可以指消息、文本、图像、视频、音频、密码协议、编码体制和系统等。已经藏有嵌入对象的输出对象称为**伪装对象** (Stego Object), 因为它与载体对象无感知差别。将嵌入对象添加到载体对象中得到伪装对象的过程称为**信息嵌入** (Information Embedding), 嵌入过程中所使用的算法称为**嵌入算法** (Embedding Algorithm)。信息嵌入的逆过程, 即从伪装对象中重新获得嵌入对象的过程称为**信息提取** (Information Extracting)。在提取过程中所使用的算法称为**提取算法** (Extracting Algorithm)。执行嵌入过程和提取过程的组织或个人分别被称为**嵌入者** (Embeddor) 和**提取者** (Extractor)。

在信息隐藏系统中, 人们通常需要使用一些额外的秘密信息来控制嵌入和提取过程, 只有它的持有者才能进行操作, 这个秘密信息称为**隐藏密钥** (Stego Key)。嵌入过程的隐藏密钥称为**嵌入密钥** (Embedding Key), 提取过程的隐藏密钥称为**提取密钥** (Extracting Key)。通常嵌入密钥和提取密钥相同, 相应的信息隐藏技术称为**对称信息隐藏** (Symmetric Information Hiding), 否则称为**非对称信息隐藏** (Asymmetric Information Hiding)。

与密码术相对应, 可以把信息隐藏的研究分为**隐藏技术** (Steganography, 隐写术) 和**隐藏分析技术** (Stegoanalysis) 两部分。前者研究向载体对象中秘密添加嵌入对象的技术。后者研究如何从伪装对象中破解出嵌入信息, 或通过对伪装对象的处理达到破坏嵌入信息或阻止信息检测目的的技术。类似地, 可以称隐藏技术的实现方或研究者**为隐藏者** (Hider), 而隐藏系统的攻击方或隐藏分析技术的研究者**为隐藏分析者** (Steganalyst) 或**伪装分析者**。

需要注意, 在信息隐藏的不同分支领域中, 上述相关术语可能对应一套不同的术语。

1.2.3 信息隐藏技术的特性和要求

1. 特性或要求

信息隐藏不同于传统的密码术，因为其目的不在于限制伪装对象正常的存取，而在于保证其中的秘密信息不被侵犯和发现。因此，信息隐藏技术必须考虑正常的信息操作所造成的威胁，即要使嵌入对象（秘密信息）对正常的数据操作技术具有免疫能力。这种免疫力的关键是要使隐藏信息部分不易被正常的数据操作（如通常的信号变换操作或数据压缩）所破坏。根据信息隐藏的不同目的和技术要求，该技术存在以下特性或要求。

（1）透明性（Transparency）或不可感知性（Imperceptibility）

利用人类视觉系统或人类听觉系统特性，经过一系列隐藏处理，使伪装对象没有明显的降质现象，而嵌入对象却无法人为地看见或听见。当然，极个别应用场合可能需要使用**可见水印技术（Visible Watermarking）**。

（2）鲁棒性（Robustness）或稳健性

指不因伪装对象通过某种常用信号处理操作而导致嵌入对象丢失的能力。这里的信号处理操作包括常见信息处理（如数据压缩、低通滤波、图像增强、二次抽样、二次量化、A/D 和 D/A 转换等）、几何变换和几何失真（如裁剪、尺度拉伸、平移、旋转和扭曲等）、噪声干扰、滤波操作、打印、扫描和多重水印的重叠等。在经过这些改变后，鲁棒的信息隐藏算法应该仍能从伪装对象中提取出嵌入的秘密信息。

（3）安全性

指算法有较强抗恶意攻击能力，它必须能够承受一定程度的人为攻击，而使嵌入对象不被破坏。此外，与加密一样，信息隐藏技术最终也需要把对秘密信息的保护转化为对密钥的保护。因此，密码术中对密钥的基本要求也适用于信息隐藏技术，如必须有足够大的密钥空间等。在设计一个信息隐藏系统时，密钥的产生、发放、管理等也必须综合考虑。

（4）不可检测性（Undetectability）

指伪装对象与载体对象需具有一致的特性，如具有一致的统计噪声分布等，以便使隐藏分析者无法判断伪装对象中是否藏有嵌入对象。

（5）自恢复性（Self-recoverability）

经过某些操作或变换后，可能会使伪装对象产生较大的破坏。如果只从留下的片段数据，仍能恢复嵌入对象，而且恢复过程不需要载体对象，这就是所谓的自恢复性。当然，并不是所有应用场合都需要自恢复性。

（6）隐藏容量（Hiding Capacity）

载体中应能隐藏尽可能多的信息。事实上，如果理想地假设伪装对象不会受到任何扰动，那么人们可在载体对象中隐藏任意多的各种不同的信息而不被察觉。当然，在保证不可感知的条件下，隐藏的信息越多，鲁棒性就越差。

事实上，信息隐藏的三个最主要因素：透明性、鲁棒性和隐藏容量，很难同时达到最优，必须根据不同的应用场景有所侧重。整个信息隐藏系统的性能将是这三个主要因素和其他因素平衡的结果。由于隐写术的研究重点是如何实现信息伪装，透明性和隐藏容量必须优先保证。而数字水印技术则需考虑鲁棒性的要求，以对抗各种可能的攻击。

2. 信息隐藏系统的攻击

对传统的信息加密系统的攻击是为了恢复或篡改秘密消息。而对信息隐藏系统的攻击除了提取和篡改秘密消息这两种形式外，还包括检测秘密信息的存在性。我们知道对

密码系统的攻击条件包括**唯密文** (Ciphertext-only Attack)、**已知明文** (Known Plaintext Attack)、**选择明文** (Chosen Plaintext Attack) 和 **自适应选择明文** (Adaptive Chosen Plaintext Attack) 四种。同样, 对信息隐藏系统也有一些相对应的攻击条件。

(1) **唯伪装对象攻击** (Stego Object-only Attack)

只有伪装对象可用于分析。

(2) **已知载体对象攻击** (Known Cover Object Attack)

可利用载体对象和伪装对象。

(3) **已知秘密消息攻击** (Known Secret Message Attack)

在某一点, 隐藏的秘密消息可能为攻击者所知。分析伪装对象, 寻找与隐藏的秘密消息相对应的模式可用于对系统的攻击。即使拥有秘密消息, 这也是很困难的, 其难度甚至等同于唯伪装对象攻击。

(4) **选择秘密消息攻击** (Chosen Secret Message Attack)

隐藏分析者用隐藏工具或算法从一个选择的秘密消息产生伪装对象。这种攻击的目标是确定伪装对象中相应的模式, 这些模式可能揭示所使用的特定的隐藏工具或算法。

1.3 信息隐藏的模型

在研究信息隐藏技术的过程中, 人们提出了不同的模型来解释信息隐藏。这些模型大致可以分为囚徒模型、通用模型、通信模型、广义模型和不对称信息空间模型五种。其中, 囚徒模型是一种经典的信息隐藏模型, 通用模型给出了利用载体实现嵌入式信息隐藏的算法流程, 通信模型从通信或信息论的角度对信息隐藏模型和隐藏容量进行研究, 广义模型从信息隐藏的机理上解释了信息隐藏存在的基础, 不对称信息空间模型从攻防双方的角度出发扩展了广义模型。下面简要介绍这些模型。

1.3.1 囚徒模型

囚徒模型是经典的信息隐藏模型 (确切的说是阈下信道模型, 详细内容参见第 6 章 6.2 节), 如图 1.1 所示。该模型以 Simmons 于 1983 年提出的“囚犯问题”作为背景。Alice 和 Bob 因犯罪被逮捕并关押在不同的囚室内。他们想策划逃跑, 不幸的是两人之间的所有通信都要在**看守者** (Warden) Wendy 的监视下进行。Wendy 不允许他们进行加密通信, 并且她如果发觉任何可疑的通信就会把他俩都送入单独的囚室并禁止任何信息交换。因此, 双方必须以隐秘的方式通信, 以免引起 Wendy 的疑心。为此, 他们不得不建立一个阈下信道。一个实用办法是将有用信息藏在某种看似普通的信息中。例如, Bob 可以画一幅画, 描绘一头蓝色奶牛躺在绿色草地上, 然后将这幅画送给 Alice。当然 Wendy 不知道画中各对象的颜色会传递信息。这里, 假设 Alice 和 Bob 在其囚室中能使用计算机系统 (对实际的囚犯可能办不到), 并能用不同的格式交换信息 (如文字、数字图像或数字音频等)。在该方案中, Alice 和 Bob 分别对应信息隐藏系统中的嵌入者和提取者, 而**看守者** Wendy 则作为隐藏分析者, 即攻击者, 存在于信息隐藏系统的信道当中。嵌入者 Alice 使用一个信息隐藏系统来向提取者 Bob 传输秘密信息, 使**看守者** Wendy 无法发现伪装对象中含有秘密信息。由于 Wendy 可能对伪装对象进行修改, 然后将修改后的伪装对象转交给 Bob。因此, 为了让秘密信息仍能传给 Bob, 该系统应该对小修改具有鲁棒性。

上述囚徒模型强调的是如何保证阈下信道的实现, 并没有提到实现信息隐藏的原

理，因此确切地说是描述了一种隐蔽通信的模型，而不是信息隐藏技术的模型；另一方面，隐蔽通信只是信息隐藏技术的一种应用，因此即使在概念上，该模型的描述也不能涵盖全部信息隐藏技术。

1.3.2 通用模型

囚徒模型概括了从古至今大多数信息隐藏技术的主要思想，其中 Alice 和 Bob 之间的隐蔽通信过程可以用更为一般的通用模型来描述。通用模型是用来描述信息隐藏嵌入流程的系统结构图。实际上，根据 1.2.2 节的术语介绍，就可得到如图 1.2 所示的信息隐藏系统的一般通用模型。

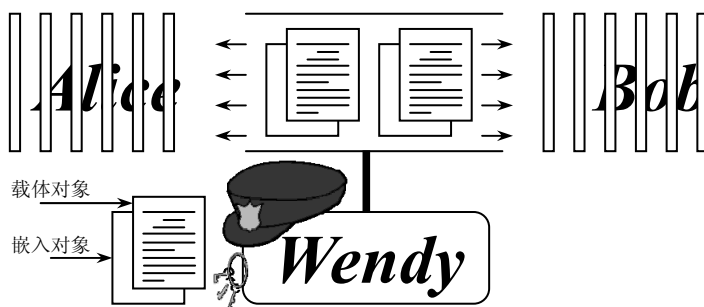


图 1.1 Simmons 提出的囚徒模型

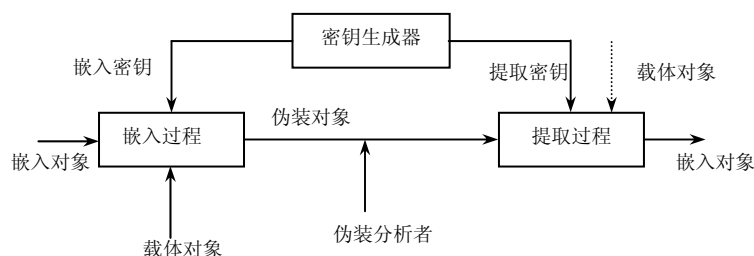


图 1.2 信息隐藏系统的一般通用模型

文献[7]给出的信息隐藏通用模型是相对比较完备和典型的，如图 1.3 所示。该模型称待隐藏的信息 m 为**秘密消息**（对应图 1.2 中的嵌入对象），它可以是版权信息或秘密数据，也可以是一个序列号。一般来说，信息隐藏系统都会将秘密消息进行预处理后嵌入，为了区分预处理前后消息的不同状态，称未预处理前的消息为明文消息 $m_p=m$ ，称经过密钥 k_{1E} 进行加密预处理的待嵌入的消息为密文消息 m_c ，而用来搭载秘密消息的载体则称为**原始载体**（Original Carrier） c （对应图 1.2 中的载体对象），如图像、声音或视频片段等。隐藏过程一般由嵌入密钥 k_{2E} 来控制，即通过嵌入算法将秘密消息隐藏于公开信息中，而**含密载体**（Stego Carrier） s （隐藏有秘密消息的公开载体，对应图 1.2 中的伪装对象）则通过公开的通信信道传递，然后提取者利用提取密钥 k_{2D} 和解密密钥 k_{1D} 从含密载体 s 中提取出秘密消息 m'_p 。其中，含密载体在信道中传输时，可能会受到以检测秘密信息存在为主的被动攻击者的攻击，也可能受到信道噪声的干扰和各种形式的主动攻击。隐藏者由于隐藏能力的限制或为了麻痹攻击者，隐藏者可能发送含密载体，也可能发送未嵌入秘密消息的原始载体。由图 1.3 可见，通用模型主要由两个模块组成：① 信息嵌入模块，利用密钥来实现秘密消息的隐藏。② 秘密消息提取模块，利用密钥从含密载体中恢复出秘密消息。在密钥

未知的前提下, 第三者很难从含密载体中发现甚至得到或删除秘密消息。在消息的提取过程中, 提取者可能用到原始载体, 也可能不需要原始载体。对于前者, 称该信息隐藏系统为**非盲隐藏** (Non-blind Information Hiding) 系统; 对于后者, 则称为**盲隐藏** (Blind Information Hiding) 系统。可见, 该模型很好地解释了利用载体实现嵌入式信息隐藏的算法流程, 但是它不能很好地解释古典信息隐藏方法, 没有明确地说明信息隐藏的实现机理。

1.3.3 通信模型

从本质上说, 信息隐藏是一种通信, 即在嵌入者和提取者之间传递秘密消息。在数字水印技术背景下, Cox 提出了最早的三种信息隐藏通信模型^[5,8]。这三种模型之间的差异在于如何将原始载体 c 融入到传统的通信模型中。在第一种基本模型中, 将原始载体完全看作噪声。在第二种模型中, 原始载体仍然被看作噪声, 但这个噪声作为**附加信息** (Side Information, **边信息**) 输入到信道编码器中。在第三个模型中, 原始载体不看成噪声, 而看成第二种信息, 该信息和秘密消息一起以多路复用的形式进行传输。这三种通信模型具有明显的缺陷: ① 它没有考虑信息隐藏的特点, 即信息隐藏是利用人类感知模型在原始载体的最不引人注意的地方嵌入秘密消息, 同时原始载体不可能完全看成噪声; ② 没有考虑鲁棒性, 所谓鲁棒性就是使得嵌入的秘密消息具有冗余, 对信号处理、有损压缩和几何攻击等具有一定的抵抗性。关于这三种模型的具体介绍参见第3章的3.3.2节, 在此不再赘述。

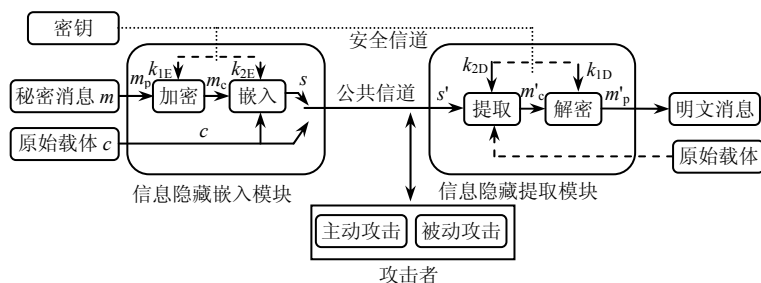


图 1.3 文献[7]给出的信息隐藏通用模型

1.3.4 广义模型

上面提到的信息隐藏模型依赖于载体和嵌入式的数学方法。载体和嵌入式的数学方法虽然保证了信息隐藏技术的快速发展, 却也限定了信息隐藏的思路, 使得其本质并没有得到扩展, 而是越来越向密码术靠拢, 甚至于一些文献开始将信息隐藏归类于密码术。然而, 信息隐藏与密码术在定义上具有本质区别。另外, 从信息隐藏的目的来看, 信息隐藏的关键在于隐藏信息的存在, 与利用怎样的载体和方法无关。为此, 林代茂等人通过对记录——感知系统的描述, 从信息隐藏的机理上解释了信息隐藏存在的基础, 提出了广义信息隐藏模型^[9]。该模型认为感知信息是一个多维矢量, 它们构成了多维信息空间 V , 而且定义在接收到的信息中, 不可感知的部分叫做**感知冗余**; 接收、解析来自外部信息的系统叫做信息的**感知系统**; 记录信息的器官或设备统称为**记录系统**。由于记录系统所能记录的信息范围 (即**记录空间** $V_r \subset V$) 与感知系统所能感知信息的范围 (即**感知空间** $V_d \subset V$) 并非全等或包含关系, 如图 1.4 所示。因此, 构造特定映射将原本可感知的信息映射到不可感知空间进行记录, 从而实现信息隐藏。该模型认为, 只要存在记录不可感知空间 ($V_r \cap \bar{V}_d$), 信息隐藏就有实现可能。下面简要介绍该模型的一些思想。

秘密消息 m 本来是可以感知的，即有

$$m \in V_d \quad (1.1)$$

但是，如果能够构造一种映射 f ，满足

$$m' = f(m), \quad m' \in \bar{V}_d \cap V_r \quad (1.2)$$

使得 m 能够从子空间 V_d 映射到子空间 $V_r \cap \bar{V}_d$ 中的 m' ，它就不能被感知了，其中 \bar{V}_d 是 V_d 的补。只要存在逆过程 f' 可以使被隐藏的信息重新回到 V_d 而被感知，即

$$m = f'\{f(m)\} \quad (1.3)$$

这样就实现了信息隐藏。以图像信息隐藏为例，人们总是设法把秘密消息 m 嵌入到载体图像 c 中，使得伪装图像 s 尽可能地接近 c 。这里，原始载体 c 就是信息记录系统，肉眼观看和各种检测方法是信息感知系统，信息隐藏的过程就是把秘密消息 m 从 V_d 子空间映射到 $V_r \cap \bar{V}_d^a$ （表示 V_r 子空间中攻击者不可感知的部分），使得伪装图像 s 尽可能接近原始图像 c ，而且满足 $m' \in V_d^m$ ，其中上标 a 和 m 分别表示感知主体是攻击者和信息隐藏的实现者。值得注意的是，从广义模型的信息隐藏机理可知，只要设法把 m 转移到可记录但不可感知的区域就可以实现信息隐藏，而并非一定要嵌入到某载体对象 c 中，所以，载体对象的存在不是信息隐藏的必要条件。

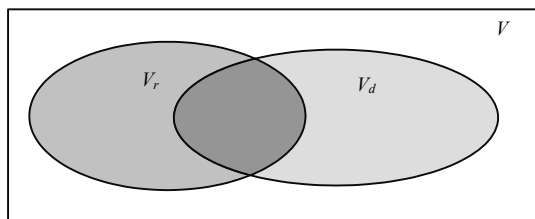


图 1.4 信息隐藏系统的广义模型

广义模型从信息隐藏的机理上解释了信息隐藏存在的基础，但未能区分参与信息隐藏的各个角色。因此，不能对隐藏过程中各个角色的行为进行区别描述，掩盖了其中所包含的一些信息隐藏特性。

1.3.5 不对称信息空间模型

为了进一步完善广义模型，李欣从攻防双方的角度提出了信息隐藏的空间模型^[10]。该模型认为“信息不对称是实现信息隐藏的基础”，定义了感知和记录空间的不对称信息空间：感知差值空间和记录差值空间。在此基础上，扩展了广义模型，将信息空间划分对于攻击者而言的 4 个部分。

(1) 记录可感知空间 A_0 。该空间内的信息可以完全被攻击者获知。因此，该空间不具备隐藏信息的能力。

(2) 感知未记录空间 A_1 。该空间内的信息为攻击者不可获取的信息。虽然其具有对该信息的感知解析能力，但该信息并没有被攻击者获取，因而不可解析该信息。

(3) 记录不可感知空间 A_2 。该空间内的信息为攻击者不可感知的信息，攻击者虽然获取了相关的记录信息，但其不具备感知能力。该空间也是现在一般算法最常利用的信息隐藏空间。

(4) 未记录不可感知空间 A_3 。该空间内的信息是攻击者既不可获知又不可感知的信息。在具体算法设计上存在一定难度。

可以看出, 除记录可感知空间 A_0 外, 其他 3 个子空间均可以被利用来实现信息隐藏。同时, 通过分析各个子空间的特性得到以下结论。

(1) “冗余”不是信息隐藏的必要条件。“记录差值空间”的存在明确的表明了隐藏不必依赖于“冗余”, 也确定了可不利用“冗余”实现信息隐藏的信息空间, 在一定程度上打破了“冗余”对信息隐藏的限制。

(2) 感知未被记录的信息可以实现隐藏。从而可以合理解释大多信息隐藏模型不能很好解释的一些基于感知未记录信息的隐藏空间的隐藏实例。

实际上, 攻防双方的记录系统、感知系统间存在的差异使信息隐藏成为可能。显然, 若隐藏某个消息 m , 必然要将该消息 m 进行变换, 使攻击方记录系统所记录的信息与 m 不同。记映射关系为 f , 攻击者记录系统记录的信息为 m' , 根据 m' 与 m 的关系将 m' 划分为不完整信息和错误信息, 即

不完整信息变换定义为

$$f_1: M \rightarrow M'_1 = M - \bar{M}'_1, \quad \bar{M}'_1 \subset M \quad (1.4)$$

其中, M 表示所有 m 的集合, M'_1 代表所有 m' 的集合。而错误信息变换定义为

$$f_2: M \rightarrow M'_2, \quad M'_2 \not\subset M \quad (1.5)$$

其中, $M'_2 \cap M = \emptyset$ 时, 称 M'_2 中的 m' 为 M 中的 m 的完全错误信息。由此, 攻击者在获得 m' 后不能解析原始秘密信息的原因也分为如下两类。

(1) 由于攻防双方记录系统的差异使得攻击方不能获得 \bar{M}'_1 , 因而 \bar{M}'_1 不能完整描述原始秘密信息。

(2) 由于攻防双方感知系统的差异使得攻击方不能解析 M'_2 。

记 Ω_s 表示隐藏空间, Ω_{rA} 和 Ω_{rD} 分别表示攻防双方的记录空间, Ω_{dA} 和 Ω_{dD} 分别表示攻防双方的感知空间。其中, 隐藏方的记录空间和感知空间可以认为是一致的。则定义双方的记录差值空间 Ω_{re} 和感知差值空间 Ω_{de} 如下:

$$\Omega_{re} = \Omega_{rD} - \Omega_{rA} \quad (1.6)$$

$$\Omega_{de} = \Omega_{dD} - \Omega_{dA} = \Omega_{rD} - \Omega_{dA} \quad (1.7)$$

$$\Omega_s = (\Omega_{rD} - \Omega_{rA}) \cup (\Omega_{dD} - \Omega_{dA}) = \Omega_{rD} - \Omega_{rA} \cap \Omega_{dA} \quad (1.8)$$

其中, 记录差值空间 Ω_{re} 内的信息为攻防双方记录系统所记录的信息的差值。感知差值空间内的信息 Ω_{de} 为攻防双方感知系统所能感知的信息的差值。秘密信息集合 M 对应的不完整信息集合 \bar{M}'_1 记录在 Ω_{re} 中, 错误信息集合 M'_2 记录在 Ω_{de} 中, 即

$$\bar{M}'_1 \in \Omega_{rA}, \quad M'_1 = f_1(M) \text{ and } \bar{M}'_1 = M - M'_1 \quad (1.9)$$

$$M'_2 \in \Omega_{dA}, \quad M'_2 = f_2(M) \quad (1.10)$$

由此可见, 若隐藏方对秘密信息 m 进行隐藏处理, 则为保证信息不被攻击者获取和发现, 其关键信息必然要映射到 $\Omega_s = \Omega_{re} \cup \Omega_{de}$ 中。综上, 基于攻防双方行为的信息隐藏模型将信息空间做如图 1.5 所示的划分。

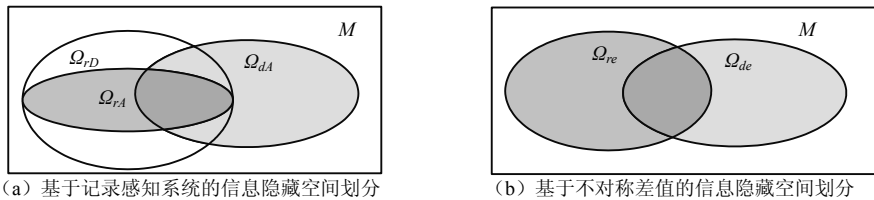


图 1.5 信息隐藏空间划分模型

1.4 信息隐藏的研究分支

信息隐藏是一门新兴的交叉学科，在计算机、通信、保密学等领域有着广阔的应用前景。信息隐藏的研究涉及密码术、图像处理、模式识别、数学和计算机科学等领域。按隐藏的应用目的和载体对象不同，信息隐藏可分为许多分支领域，如图 1.6 所示。下面对信息隐藏的几个主要分支作简单介绍。

1.4.1 隐写术

信息安全的研究不仅包括密码术的研究，还包括信道安全的研究，其实质在于隐藏信息的存在。信息隐藏应用于这一领域的一个重要分支就是**隐写术**。“隐写术”一词源于希腊词语“*στεγανω*”，其字面意思是“掩饰性地写”，即将秘密消息隐藏在其他消息中。隐写术的主要目的是将重要的信息隐藏起来，以便不引人注意地传输和存储。因此，在隐写系统中，嵌入对象是秘密消息，即通过隐写手段保护的主体，而载体对象可以是任何能够达到隐蔽传输目的的载体数据。通常情况下，选择载体对象时需要考虑**隐写容量**（Stego Capacity）的大小和隐写结果的不可感知性这两方面因素。在隐写信道中存在的隐藏分析者常常被称为**看守者**，这是因为人们通常将 Simmons 提出的“囚犯问题”作为隐写系统的通用模型。如果看守者能对流经他的信息流进行改动和处理，则称之为**主动看守者**（Active Warden），否则称之为**被动看守者**（Passive Warden）。隐写术大致可分为两种，一种是将需要秘密传递的消息记录下来，然后通过其他媒介发送出去的技术，即“**技术隐写术**（Technical Steganography）”；另一种是将记录这个行为本身隐藏起来的技术，消息由隐藏地“写”的语言或语言形式所组成，即**语义隐写术**（Linguistic Steganography）。例如，感应墨水技术、缩小影像术和扩频通信技术都是技术隐写术，而语义隐写术包括**符号码**（Semagram）、**隐语**（Open Code）及**虚字密码**（Null Cipher）等。根据隐写的思想方法，可将隐写术分为基于隐匿安全的、基于伪装的、基于隐藏嵌入位置的、基于扩散待嵌入信息的和特殊环境下的隐写术等。有关隐写术的详细介绍见第 2 章。

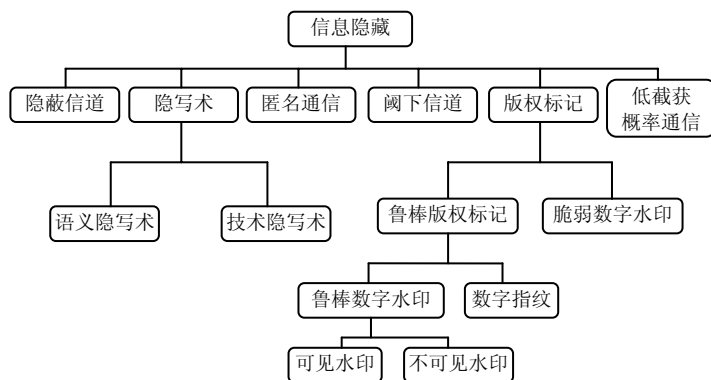


图 1.6 信息隐藏的主要研究分支

隐写术和密码术都是一门古老的学科，两者都为消息收发双方提供机密性、完整性、可鉴别、抗抵赖的解决方案。密码术是对记录进行保密，隐写术是对记录进行隐蔽。密码术以公开方式传递密文，不隐蔽秘密信息本身的存在，容易引起攻击者注意。而隐写术则以秘密的方式传递**明文**（Plaintext），隐蔽消息的存在，从而对消息内容隐蔽的要求就减少了，但为增加保密性，通常对嵌入对象先加密再隐藏。因此，隐写术是密码术的很好补充。

1.4.2 版权标记

随着计算机和网络的飞速发展,人们的许多创作和成果都以数字形式进行存储和发布。然而,数字作品极易被非法复制、伪造或篡改,使得很多版权所有者不愿利用网络公开其作品,从而阻碍其自身发展。目前,数字作品的版权保护不仅仅是立法问题,也是一个很重要的技术难题。从技术上看,数字媒体版权信息的嵌入和检测问题,是数字作品版权保护的两个关键问题,它综合了传统密码术的认证和鉴别问题的特点,又加入了鲁棒性要求。版权保护信息必须与被保护的数据密切结合,版权保护信息的鉴别过程必须具有抗干扰能力。在这种情况下,向数字作品中加入不易察觉的但可以判定区分的标记是数字作品版权保护的一种有效技术手段,这种技术统称**版权标记技术**(Copyright Masking)。这种标记可以是一个隐藏的商标、序列码、或是防止未授权用户直接复制的方法。根据标记内容和目的的不同可将版权标记技术分为**数字水印技术**和**数字指纹技术**两大类。在数字水印系统中,**数字水印**(Digital Watermark)所代表的是作者的身份,该信息必须经过注册获得管理机构的认可,多个作品可嵌入相同**水印**(Watermark);而在数字指纹系统中,**数字指纹**(Digital Fingerprint)所代表的是购买者的身份,同一作品需要嵌入不同的**指纹**(Fingerprint)对不同的购买者进行区分。需要注意,本书将版权标记技术看作广义上的数字水印技术。在数字水印/数字指纹系统中,隐藏分析者通常被称为**盗版者**(Pirate)/**叛逆者**(Traitor)。

1.4.3 隐蔽信道

所谓**隐蔽信道**,就是指在计算机安全技术中,一种允许某个进程在违反安全规则的状态中传递信息的信道,或者说是一种允许违背合法的安全策略的方式进行操作系统**进程间通信**(Inter-Process Communication, IPC)的通道。隐蔽信道出现的根本原因是计算机系统的安全机制本身。对于一个系统整体来说,采用统一的安全性是不切实际的。通常情况下,人们将系统的某个部分分离开,专门处理敏感信息,有较高安全性,而系统的其余部分则处理敏感性相对较低的信息。在系统中,某个安全级别的部分只能对相同安全级别或者更低安全级别的数据进行读写,而对高安全级别的数据只能读或者根本不能访问,这就是一种多级安全的概念。**多级安全系统**(Multilevel Secure System)是一类包含不同等级敏感信息的系统,它既可供那些确有必要且具有不同安全许可和已知需要的用户同时访问,又能阻止非法用户去访问其无权过问的信息。这种系统存在一定的隐患。假设系统的某一个不可信任的部分 P 具有高的安全级别 H ,它可以通过某种方式影响该系统,使得自己的操作可以被比它低级或级别相当的部分 L 看到。那么, P 就能够利用这样的状况建立一条从 H 到 L 的通道,把 H 的数据进行编码,并通过该通道传送给 L 。由于 P 部分和 L 部分都不够安全,那么诸如特洛伊木马之类的病毒代码就能经由该通道植入到这两部分中的任何一个。现代隐蔽信道的重要参数是带宽。它分为**隐蔽存储信道**(Covert Storage Channel)和**隐蔽时间信道**(Covert Timing Channel)两种。隐蔽存储信道包含由一个进程直接或间接写一个存储地址,而由另一个进程直接或间接读一个存储地址的隐蔽信道。在隐蔽时间信道上,一个进程通过调整自身对系统资源(如 CPU 时间)的使用,向另一进程发送信息。这种处理又影响了第二个进程观测到的实际响应时间。隐蔽信道中的概念与隐写术中的许多概念类似,但在隐蔽信道中,载体对象是一个系统的整个运行过程,而不是具体的信息媒介。系统开发者应彻底搜索隐蔽信道,并根

据实际测量或工程估算确定每一个被标识信道的最大带宽。隐蔽信道是不可能完全阻断的。尽管如此，隐蔽信道研究人员仍在寻找检测和控制隐蔽信道更有力的方法。

1.4.4 阈下信道

Simmons 于 1978 年发现苏美两国签署的 SALT II 条约的监督协议中存在致命的漏洞，从而提出阈下信道的概念。阈下信道也叫潜信道，它是指在公开信道（Overt Channel）中所建立的一种实现隐蔽通信的信道。密码协议中的阈下信道是指被用来传输秘密消息的各种编码体制和密码协议中所采取的数学结构。经典密码体制中不存在阈下信道，DES（Data Encryption Standard）加密方案中也不存在阈下信道，因为明文块和其对密文块的大小相同。但是，在大多数基于公钥体制的数字签名方案中，明文与数字签名并不是一一对应的，这是由于会话密钥具有可选择性，从而对同一个消息可产生多个数字签名。虽然这并不影响对签名的验证，即对于多个不同的签名，只要私钥相同，那么验证者就可通过适当计算来验证签名的有效性，但这就为阈下信道的存在提供了条件，阈下收方（Subliminal Receiver）可以根据这些不同的数字签名获取公开收方（Overt Receiver）无法得到的阈下信息/消息（Subliminal Information/Message）。研究表明，绝大多数数字签名方案都可包含阈下信道的通信，其最大特点是阈下信息对数字签名和验证的过程无任何影响，这正是其隐蔽性所在。即使监视者知道要寻找的内容，也无法发现信道的使用和获取正在传送的阈下消息，因为阈下信道的特性决定其安全保密性要么是无条件的，要么是计算上不可破的。

众所周知，如果发送方能够影响和控制载体对象，那么就可采用隐写术来进行隐蔽通信。但是，如果发送方无法影响或控制载体对象的产生，而检查者可能知道载体信息的任何细节，甚至可以控制该公开信道，并且任何信息到达接收方前都必须由检查者进行严格检查认证，那么阈下信道将比隐写术更加安全。另一方面，阈下信道可以看作隐蔽信道的特例。但是，阈下信道通常在密码协议执行过程中建立，它以密码协议为载体对象，涉及信息论和计算复杂度理论，不是工程实现上的方法；而隐蔽信道可用任何技术建立，通常涉及实现方面的技术。

由于阈下信道存在于密码协议中，故它所带来安全性隐患非常巨大。阈下信道多年来一直是国际上诸多机要部门研究的重点课题之一，关于阈下信道的利用和反利用是一种处于更高层次上的高技术高智力较量。阈下信道在国家安全方面的应用价值很大。如果采用全球性标准，那么世界上任何地方的用户检查点都能即时检查出数字证件上的信息完整性，并能确定持证人是否为合法持证人。将来可以在数字签名的数字证件中建立阈下信道，把持证人是否为恐怖主义分子、毒品贩、走私犯或重罪犯等情况告诉发证国的海关人员，以及向金融机构、商业实体透露持证人的信用评价和支付史情况，而仅检查公开信息的人无法看到此类阈下信息，持证人自己也无法获得和修改这些阈下信息。

1.4.5 低截获概率通信

在现代战争中，即使内容已被加密，敌人也会从发现一个信号而迅速发起对发送者的攻击。因此，截获与反截获是机要通信战线上的重要斗争之一。正是由于这种军事上的刺激，低截获概率通信的研究已经越来越引起各国政府和军队的重视，成为现代通信的重大课题。低截获概率通信，顾名思义就是使信号被截获的概率降低的通信技术，其载体对象是整个通信频带。它主要包括扩展频谱通信技术和流星猝发通信技术。

扩展频谱通信是 20 世纪 50 年代开始研究的一种通信技术,最初用于军事领域。1985 年后用于商业系统,从此该技术在全球定位系统以及数字移动电话等领域得到广泛应用和快速发展。扩展频谱通信是将待传送的信息数据用伪随机编码调制,实现频谱扩展后再传输,而接收端则采用同样的编码进行解码和相关处理,恢复原始信息数据。这样,扩频通信把集中在较窄频段的待发送信息展宽到较宽频带,并可以在很低的信噪比下传送信息,在不知道随机编码机制的情况下,截获低功率谱密度的扩频通信信息是一件很困难的事情。该技术的两个重要特点是伪随机编码调制与信号相关处理。它具有抗干扰、保密性、多址复用和任意选址等优点,得到广泛应用。

微流星体电离轨迹进入大气层会反射无线电波,特别是在低超高频带宽,若进行适当导向,可以用此建立远程通信连接。由于流星轨迹会迅速散射,信号强度会迅速衰减,流星猝发通信的这一间歇特性以及每一条轨迹相对较小的覆盖区域,使得它们本质上难以被监视,可用于低截获概率通信。流星猝发通信技术是一门交叉学科,它覆盖了从流星轨迹本身的物理研究,不同编码以及其他用于最大化可用信道容量的技术。它具备造价低、可靠、生存力强、远程等特点,在商业、军事、天气预报、远距离遥测、车辆跟踪、双向通信等方面有很大应用价值。

1.4.6 匿名通信

匿名通信(Anonymous Communication)就是寻找各种途径来隐藏信息的发送者和接收者,其中使用的主要技术包括匿名重发和网络代理等技术。在网络通信中,跟踪敌手的数据包,进行业务量分析,以及判断谁和谁在通信,也是收集谍报信息的一个重要手段,而采用匿名通信技术就是为了保护通信信道不被别人窃听和进行业务量分析。这种技术提供一种基于 TCP/IP 协议的匿名连接,从数据流中除去用户的标识信息,即用该技术建立连接时,并不是直接连到目的机器的相应数据库,而是通过多层代理服务器,层层传递后到达目的地址,每层路由器只能识别最临近的一层路由器,第一层路由器对本次连接进行多层加密,以后每经过一层路由器,除去一层加密,最后到达的是明文,这样每层路由器处理的数据都不同,使敌手无法跟踪。连接终止后,各层路由器清除信息。这有点类似于地下工作者的单线联络,每个人只知道与前后哪两个人接头,而对自己所传的消息最初从哪儿来、最终到哪儿去一概不知。从应用的角度讲,比如你在网上购物,不想让中途窃听者知道你买了什么东西、你是谁或者你访问某个站点,就可以用这种技术。这种技术可用于有线电话网、卫星电话网等,它不但适用于军用,而且适用于商业,还可广泛用于 E-mail、Web 浏览以及远程注册等。Web 应用强调接收方的匿名性,而电子邮件用户们更关心发送者的匿名性。此外,匿名通信技术还可用在电子选举和电子现金方案中来保证选举人或者购买者的身份不被泄漏。

1.5 信息隐藏技术的分类

除了 1.4 节中按照研究分支(或者保护对象)分类外,信息隐藏技术还可以按照其他不同的方式进行分类,比如可以按照载体对象类型分类,也可以按照密钥的对称性分类,也可以按照嵌入域分类。下面简要介绍各种分类方法。

1.5.1 按载体类型分类

按照载体类型分类,信息隐藏技术可以分为基于文本、图像、音频、视频、三维模

型和动画等各种不同媒体的信息隐藏技术。

文本信息隐藏,是通过改变文本模式或改变文本的某些基本特征来实现信息嵌入的方法,它使文档产生一定的变化,但是人的视觉对这种变化是不易察觉的。信息隐藏是充分利用载体中的冗余信息的存在来工作的。由于人类视觉和听觉系统对某些信息不敏感,图像、声音、视频文件天然地包含噪声形式的冗余,但是在文本里面隐藏信息是比较困难的。文本文件是直接对文字数据进行编码而成,因此几乎不存在数据冗余,不可能通过修改原文件的有效数据来进行信息隐藏,必须寻找那些不易引起视觉感知的方法。目前,在文本中隐藏数据主要是将信息直接编码到文本内容中去(利用语言的自然冗余性),或者将信息直接编码到文本格式中(比如调整字间距或行间距),或者利用人们通常不易察觉的标点和字体的改变等方法,具体方法将在第2章中介绍。

基于图像的信息隐藏技术是近年来信息隐藏技术中最具挑战性、最为活跃的研究课题之一,它以数字图像为掩护媒体,将需要保密的信息按照某种算法嵌入到数字图像中。图像是像素(Pixel)的集合,相邻像素点所对应的实际距离称为图像的空间分辨率。根据像素颜色信息的不同,数字图像可分为二值图像、索引图像、灰度图像以及RGB彩色图像。数字图像的最终感受者是人的眼睛,人眼感受到两幅非常相同的数字图像的像素值可以存在较大的差别。这样,依赖于**人类视觉系统(Human Visual System, HVS)**的不完整性,就为数字图像的信息隐藏提供了非常巨大的施展空间。关于图像信息隐藏的具体方法将在第2章和第3章中介绍。

音频信息隐藏技术是在音频信号中嵌入不可察觉的秘密信息,以实现版权保护、隐蔽通信等功能。音频信号作为信息隐藏载体,本身有许多不同于图像和视频的特征,具体如下。① 听觉系统虽然很灵敏,但存在时间和频率掩蔽效应,可以通过恰当的嵌入方法,来掩盖数据嵌入带来的失真。② 相对于图像和视频,音频的处理不需要大量的计算,适合实时处理,而且语音和音乐的录制也比较方便。③ 在使用电话或手机进行通信时,少量的噪音不会引起注意和不适。④ 在过去的几年中,图像是隐藏系统所偏爱的载体,然而自有报导恐怖分子用它们来传递信息以后,对图像的隐写比较警觉,相对而言,音频还是比较安全的载体。这些特征也使针对音频特点的信息隐藏研究愈发显得重要。关于音频信息隐藏的具体方法将在第2章和第3章中介绍。

视频作为信息隐藏的载体,较图像、音频等媒体具有更大的信号空间,因而可以隐藏较大容量的信息,为保密通信、版权保护、内容认证等问题提供解决方案。以视频为载体的信息隐藏技术,利用了人眼的视觉特性——分辨率与灵敏度上的局限性,在载体信号的感知冗余中嵌入秘密信息。基于视觉特性的视频信息隐藏具有重要的研究价值和广阔的应用前景,其主要难点是如何建立视觉感知模型并用以指导视频信息隐藏。关于视频信息隐藏的具体方法将在第2章和第3章中介绍。

当前的数字水印技术大都是针对静止图像、音频流和视频流这类媒体数据类型的,而对三维几何模型数据的水印技术的研究工作相对较少。随着三维扫描技术的发展,三维几何模型已经成为继音频、图像、视频后的一种新的媒体类型,被广泛应用于娱乐业和制造工业以及其他各种领域中。三维几何模型的版权保护问题也变得日益重要。三维几何模型数字水印技术作为数字水印技术的一个分支出现了。现在三维数字水印技术尚处于初期阶段,有很多问题需要解决。其中,三维数字水印攻击种类繁多,现有的三维数字水印技术通常只对其中小部分攻击方式有一定的对抗性,而其他方法的攻击则会造成水印无法提取,因此对现有的三维数字水印方法进行改进,使之能对抗更多种

类的攻击，最终达到对所有种类的攻击都有一定的鲁棒性是三维数字水印技术的一个重要的方向。

1.5.2 按密钥对称性分类

按照嵌入和提取过程是否需要密钥以及密钥的对称性，信息隐藏技术可以分为无密钥信息隐藏、**对称信息隐藏**（私钥信息隐藏）和**非对称信息隐藏**（公钥信息隐藏）。

如果一个信息隐藏系统不需要预先约定密钥，称其为无密钥信息隐藏系统。在数学上，其信息隐藏过程可描述为一个映射 $E_m: C \times M \rightarrow S$ ，这里 C 是所有可能载体的集合， M 是所有可能秘密消息的集合， S 是所有伪装对象的集合。信息提取过程也可看作一个映射 $E_x: S \rightarrow M$ ，从伪装对象中提取机密消息。在所有实用的信息隐藏系统中，载体集合 C 应选择为由一些有意义的，但表面上无关紧要的消息所组成（如所有有意义的数字图像的集合），这样通信双方在交换信息的过程中不至于引起监视者的怀疑。嵌入函数应该满足使载体对象和伪装对象在感知上是相似的这样一个条件。无密钥隐藏算法收发双方不需要预先预定密钥，但必须约定嵌入算法和提取算法且这些算法必须保密。这样一来，无密钥信息隐藏系统的安全性完全依赖于隐藏和提取算法的保密性，如果算法被泄漏，信息隐藏就无任何安全性可言。在密码术的研究中，有一个公认的设计准则就是1883年 Kerckhoffs 阐明的第一个密码系统的设计准则：密码设计者应该假设对手知道数据加密的方法，数据的安全性必须仅依赖于密钥的安全性。在密码设计时应该考虑满足 Kerckhoffs 准则。信息隐藏的安全性也同样存在这样的问题，信息隐藏系统的设计也应该考虑满足 Kerckhoffs 准则。无密钥信息隐藏系统的安全性完全建立在隐藏算法的安全性上，显然违反了 Kerckhoffs 准则，在现实中是很不安全的。

为了提高无密钥信息隐藏技术的安全性，可以将秘密信息先进行加密，再进行隐藏，使信息得到了两层保护，一个是用密码术将信息本身进行保密，另一个是用隐藏技术将信息传递的事实进行掩盖，这样就比单独使用一种方式更安全。若嵌入和提取采用相同密钥，则称其为对称信息隐藏（私钥信息隐藏），否则称为非对称信息隐藏（公钥信息隐藏）。同密码术一样，信息隐藏系统的安全性不能靠算法保密来保证。在设计安全的信息隐蔽传输算法时，应该假定信息隐藏算法是公开的，也就是说，在信道上监视的非授权者知道信息隐藏算法，他可以对用户 A 与用户 B 之间传递的每一个载体对象进行分析，用相应的信息提取算法提取机密信息。但是在他不知道伪装密钥的情况下，无法提取出有效的秘密信息，正如已知加解密算法，但不知道密钥仍然无法破译密码系统一样。一个私钥隐藏系统类似于私钥密码系统：发送者选择一个载体对象 c ，并使用伪装密钥 k 将秘密信息 m 嵌入到 c 中。伪装密钥是由发送者和接收者所共同拥有的（可以事先约定，也可以同时产生，其产生和使用方法等同于密码术中的密钥交换协议）。接收者利用手中的密钥，用提取算法就可以提取出秘密信息。不知道这个密钥的任何人都不可能得到秘密信息。私钥隐藏系统需要密钥的交换。在密码术中，总是假定通信各方都能够通过一个安全的信道来协商密钥，并且有各种密钥交换协议，以保证每次使用的密钥的安全保密性。信息隐藏系统一个独有的特点是可以直接将伪装密钥“放在”载体中传递给对方。例如，通过利用载体的某些内在特征和一个安全哈希函数，完全可以直接从载体计算出一个用于秘密通信的密钥： $k=h(\text{Feature})$ 。如果嵌入处理不改变载体的“内在特征”，接收者就能重新计算出密钥 k 。当然，这种“内在特征”必须高度依赖于载体，使得接收方能准确地从这种“内在特征”中恢复出密钥。

为了使信息隐藏的使用更加方便和安全,非对称信息隐藏(公钥信息隐藏)系统的概念已经被提出。这对信息隐藏系统的算法设计提出了更高的要求。公钥信息隐藏系统使用了公钥密码系统的概念,它需要使用两个密钥:一个公钥和一个私钥。通信各方使用约定的公钥体制,各自产生自己的公钥和私钥,将公钥存储在一个公开的数据库中,通信各方可以随时取用。私钥由通信各方自己保存,不予公开。公钥用于信息的嵌入过程,私钥用于信息的提取过程。公钥信息隐藏的协议由 Anderson 首次提出,它的方法是:用户 A 利用用户 B 的公钥来对需要保密的消息进行加密,得到一个“外观”随机的消息,将它嵌入到一个载体对象中去,嵌入方法就是替换掉载体的测量噪声。任何数字化的载体信号都存在或多或少的测量噪声,测量噪声具有“自然随机性”。如果加密后的消息可以达到近似于“自然随机性”,那么信息嵌入后不会影响载体的感官特性。这里还可以假设加密算法和嵌入函数是公开的,因此任何人都可以利用提取函数得到外观上随机的序列。但是只有接收者 B 拥有解密密钥,用解密密钥可以解出用户 A 发来的秘密信息。第三方监视者虽然也可以得到这样的随机序列,但是由于他不拥有解密密钥,无法肯定这样的随机序列是载体信号的自然噪声还是秘密信息被加密后产生的随机序列。当然,接收者 B 也无法肯定他每次接收的伪装对象中都包含有加密信息,因此他只能每次运行解密算法,试图用私钥去解密,如果伪装载体确实含有秘密信息,则解密出来的就是用户 A 发来的秘密消息。公钥信息隐藏的安全性取决于所选用的公钥密码体制的安全性。同时,还要求用公钥加密后的数据具有良好的随机性,且该随机性应与载体测量噪声的自然随机性在统计特征上是不可区分的。

1.5.3 按嵌入域分类

按照嵌入域分类,信息隐藏方法主要可分为原始域(包括空域、时域和时空域)、变换域和压缩域方法。原始域方法是指在载体的原始数据域内,根据一定规则通过直接修改原始数据值实现信息隐藏的一类方法,通常用秘密信息替换载体对象中的冗余部分,它包括空域、时域和时空域三种。空域方法是指在文本、图像、图形、三维模型和视频等载体的原始空间域内,根据一定的规则通过直接修改像素值、位置坐标或间隔大小实现信息隐藏的一类方法。例如,一种简单的空域替换方法就是用秘密信息位替换载体中的一些**最不重要位**(Least Significant Bit, LSB),只有知道隐藏信息嵌入的位置才能提取信息。此方法较为简单但鲁棒性较差。对载体的小扰动,如有损压缩都可能导致整个信息的丢失。时域方法是指在音频和视频等载体的原始时间域内,根据一定的规则通过直接修改时域采样值或采样间隔实现信息隐藏的一类方法。例如,回声隐藏方法利用人耳听觉系统的掩蔽效应,将秘密信息嵌入到人耳不可感知的区域,通过不同的延迟来代表不同的秘密信息。时空域方法主要针对视频和动画而言,根据一定规则通过同时在时域和空域直接修改原始载体实现信息隐藏的一类方法。

变换域信息隐藏算法将原始载体信号进行某种变换,如**快速傅里叶变换**(Fast Fourier Transform, FFT)、**离散余弦变换**(Discrete Cosine Transform, DCT)和**离散小波变换**(Discrete Wavelet Transform, DWT)等,获得变换域的系数,通过修改变换域系数的方法实现秘密信息的嵌入。一般低频系数的变化会影响到载体信号的感知效果,而高频系数的修改会造成鲁棒性的降低,所以许多算法通常修改中频系数来达到不可感知性和鲁棒性的折衷。一般而言,变换域方法对诸如压缩、修剪和某些图像处理等攻击的鲁棒性更强。与原始域相比,变换域方法具有以下优点。① 在变换域中的信号能量可以分

布到载体的所有采样上；② 在变换域中，人的感知系统的某些掩蔽特性可以更方便地结合到编码过程中；③ 变换域方法可与数据压缩标准如 **JPEG**（Joint Photographic Experts Group）等兼容。

随着多媒体压缩技术的研究和进步，许多载体对象通常以压缩格式文件存储，所以直接在压缩域嵌入秘密信息成为学者们关注的焦点。压缩域隐藏技术是将隐藏过程和压缩过程相结合来实现秘密信息嵌入的一种技术，它能有效地避免感知编码对隐藏的攻击。比如在 **JPEG** 图像压缩过程中对量化因子和变换域系数来嵌入秘密信息能有效地抗 **JPEG** 解压缩与再压缩的攻击。目前最流行的音频压缩域隐藏方法是基于 **MP3**（MPEG-1 or MPEG-2 Audio Layer III）格式的压缩域隐藏。目前最流行的视频压缩域隐藏方法是基于 **MPEG**（Moving Picture Experts Group）、**H.26X** 和 **AVC**（Advanced Video Coding）的压缩域隐藏。

1.5.4 其他分类方式

除了以上三种主要的分类方式，还有其他几种分类方式如下。

按秘密信息提取时是否需要原始载体对象的参与，信息隐藏技术可分为**非盲隐藏**（私有隐藏）和**盲隐藏**（公有隐藏）两类。若提取秘密信息时不需要原始载体对象的参与，则为盲隐藏。若提取秘密信息时需要原始载体信息的参与，则为非盲隐藏。显然，使用原始载体对象参与更便于检测和提取出秘密信息。但是，在数据监控和跟踪等场合，我们并不能获得原始的载体。所以，盲隐藏是学者们研究的重点。

按照可逆性进行分类，信息隐藏技术可以分为**可逆信息隐藏**和**不可逆信息隐藏**。可逆信息隐藏是指在原始载体对象中嵌入秘密信息形成伪装对象后，不但可以从伪装对象中解码出秘密信息，而且还可以复原到原始载体对象。不可逆信息隐藏只能从伪装对象中解码出秘密信息，但不能复原出原始载体对象。

按鲁棒性分类，信息隐藏技术可以分为**鲁棒信息隐藏**、**脆弱信息隐藏**和**半脆弱信息隐藏**。鲁棒信息隐藏系统是指在保证伪装对象与原始载体对象的感知相似条件下，在各种无意和恶意攻击下，秘密信息仍然不能被修改、去除的信息隐藏系统。脆弱信息隐藏系统则相反，对各种无意和恶意的攻击下，所隐藏的秘密信息都会丢失。而半脆弱信息隐藏系统是对某些攻击鲁棒而对其他攻击脆弱的信息隐藏系统。

按照要保护的对象分类，信息隐藏技术主要可以分为**隐写术**和**版权标记技术**。隐写术保护的对象是秘密信息，其目的是在不引起任何怀疑的情况下秘密传送信息，因此它的主要要求是不被检测到和**嵌入容量**（Embedding Capacity）大等。版权标记技术保护的对象是数字产品，该技术可以进一步分为数字水印技术和数字指纹技术，水印和指纹是嵌入在数字产品中的数字信号，可以是图像、文字、符号和数字等一切可以作为标识和标记的信息，其目的是进行版权保护、所有权证明、版权跟踪（数字指纹）和完整性保护等，因此它的主要要求是鲁棒性和不可感知性等。

1.6 信息隐藏技术的历史发展

要想掌握一门学科，就有必要了解它的发展历史，目的有两个方面：一方面，“以史为鉴”，避免犯前人同样的错误；另一方面，继承前人的优秀思想。下面，将信息隐藏技术的历史发展分为古代、近代和现代三个阶段分别进行说明。

1.6.1 古代信息隐藏技术

自从出现人类文化，人类就有保护信息的想法。**密码术**（Cryptography）和**隐写术**（Steganography）这两个词正式出现在 17 世纪中叶，且都来源于希腊语，其中隐写术对应的英文意思是“Covered Writing”。实际上，信息隐藏的概念，最早可以追溯到远古时期的古希腊时代。有“历史之父”之称的希腊史学家 Herodotus（公元前 484～公元前 425），撰写了一部伟大的著作《历史（The Histories）》，叙述波斯与希腊战争的始末。书中就有提到几个与信息隐藏有关的经典故事，例如：公元前 490 年，波斯帝国的 Darius 一心想想要占领希腊，于是希腊各殖民地打算联合起来背叛波斯，当时 Miletus 的统治者 Histiaeus 不顾众人的反对，将这件事报告给 Darius 知道。Darius 大为赞赏，于是封他为 Miletus 的国王，并且让他的女婿 Aristagoras 出任 Miletus 的统治者。只是 Darius 的弟弟担心 Histiaeus 统一希腊城邦后，会反过来与波斯为敌，于是劝 Darius 打消这个念头。因此 Darius 将 Histiaeus 召回，并且软禁在波斯的国都 Susa 城。当 Histiaeus 被软禁时，为了联络女婿 Aristagoras，他找了一名相当信任的奴隶，将奴隶的头发全部剪光，并且在头皮上刺上要给 Aristagoras 的信息，等到奴隶长出头发后，再派奴隶回 Miletus。这样即使 Darius 捉到这名奴隶，也不知道奴隶头发下暗藏玄机。之后，公元前 480 年，前斯巴达国王 Demaratus 被驱逐至 Susa 城，Darius 的儿子波斯国王 Xerxes 收留了他。虽然 Demaratus 被驱逐出境，但他仍然很效忠希腊，当他知道 Xerxes 有意突袭希腊时，决定写信将 Xerxes 的入侵计划告诉斯巴达现任国王 Leonidas。但是，城里到处都有卫兵看守，如何才能确保信的内容不被发现呢？聪明的 Demaratus 拿了一个木制蜡板，刮除上层的蜡后，将 Xerxes 的侵略计划刻在木板上，再将木板以蜡封住，由于蜡板的外观还是跟之前一样，因此可以顺利躲过卫兵的检查。Demaratus 的木制蜡板后来被 Leonidas 的妻子 Gorge 破解。希腊人在万全准备的情况下，让波斯的军队在一天之内就瓦解了。公元前 400 年，斯巴达的 Lysander 将军从密使手中截获一根名为斯巴达密码棒（Scytale）的锥形木棒及一张皮条，如图 1.7 所示，Lysander 将皮条沿着固定角度螺旋形缠绕在木棒上，结果可以看到写在木棒上的信息，得知波兰帝国的 Pharnabazus 准备要攻击他，这个密讯使得 Lysander 有足够的时间做准备，以对抗 Pharnabazus 的军队。

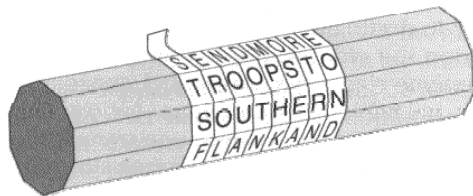


图 1.7 斯巴达密码棒

除了 Herodotus 提到的例子外，还有许许多多的古代信息隐藏例子，例如：16 世纪时，苏格兰的玛丽女王秘密策划准备暗杀英格兰的伊丽莎白女王，她将秘密信件密封在酒桶中空的塞子中，再传给同谋者。只是后来被同谋者出卖，信件密码被破解，因而被判处死刑。古代中国也有类似的案例，一名将军将机密信息写在丝绸上面，再将丝绸揉成小圆球，并且以蜡将小圆球封起来，接着命令传令兵将小圆球吞入肚中，以达到机密传递的目的。还有，用信鸽传递秘密消息这一古老的方法至今仍在被使用。1850 年，路透社的创刊人 Reuter 就利用鸽子传递信息，从而战胜了他的竞争者。

更快速、更常见的一种隐写术是将信息隐藏在看似普通的文字当中，如在书中所选

字母下面扎孔，它们可以拼出完整的信息内容。在英国，这种做法曾盛行一时，当时寄信在英国价格不菲，但邮局寄送报纸却是免费的，于是，一些家庭通过在报纸扎上肉眼几乎看不到的针孔来传递信息。另一种方法是，如果你在原始文本上写字，你可以在文字本身隐藏信息内容。此类密码传递信息的方式有时也存在令人拍案叫绝的案例，例如詹姆斯王钦定版《圣经》诗篇 46 中第 46 个字是“shake”。从诗篇 46 末尾向后数的第 46 个字是“spear”，组合起来就是 shakespear（莎士比亚）。詹姆斯王版本的《圣经》于 1610 年出版，那一年是英国大文豪威廉·莎士比亚的 46 岁生日。当时，负责翻译这部《圣经》的英格兰知名学者均被要求在书中增加插图，以向一贯遭到轻视的英格兰国王詹姆斯一世表示尊重。显然，有些学者决定将他们对一个真正敬重之人的钦佩之情隐藏在书中。历史上诸如此类的隐写方法还有多种。17 世纪，英国的 Wilkins（1614～1672）是资料记载中最早使用隐写墨水进行秘密通信的人。在《The Davinci Code》一书中，叙述了罗浮宫馆长 Jacques Saunier 在死前利用隐形墨水在地上及圣母画像后面，留下信息给孙女 Sophie Neveu 和哈佛大学教授 Robert Lagdon。这种隐形墨水只有用红外线、紫外线灯光照射，并且在黑的环境下才会显现，这种墨水常被用来标记艺术品是否已经修复的工具。我们可简单的利用一些有机液体，例如牛奶、橘子汁、醋、尿等，来制作隐形墨水，因为这些液体含有丰富的碳元素，当液体干燥后再利用火加热，就会使得液体炭化，显现出字迹。早期的隐写墨水是由易于获得的有机物（如牛奶、果汁或尿）制成，加热后颜色就会变暗从而显现出来。

在中国古代，人们曾经使用挖有若干小孔的纸模板盖在信件上，从中取出秘密传递的消息，而信件的全文则是为打掩护用的。其实，由于中国古代文字狱盛行，所以民间流传的“藏头诗”成为人们传递某些特定信息的载体工具。当然，“藏头诗”并不一定就是中国古代劳动人民为了反对当时的封建政府文字狱而发明的，但是“藏头诗”的存在却也恰恰证明了人们对信息隐藏的研究与认识。例如《水浒传》中梁山好汉们为了拉卢俊义入伙，“智多星”吴用和宋江便生出一段“吴用智赚玉麒麟”的故事来，利用卢俊义正为躲避“血光之灾”的惶恐心理，口占四句卦歌：芦花丛中一扁舟，俊杰俄从此地游。义士若能知此理，反躬难逃可无忧。此诗每句开头一个字连起来就暗藏“卢俊义反”四字，广为传播。结果，成了官府治罪的证据，终于把卢俊义“逼”上了梁山。再比如明朝大学问家徐渭（字文长）游西湖，面对平湖秋月胜景，即席写下了七绝一首：平湖一色万顷秋，湖光渺渺水长流。秋月圆圆世间少，月好四时最宜秋。此诗每句开头一个字连起来就暗藏了“平湖秋月”四字。其实，在中国古代，还有一种非常流行的“隐写术”——谜语。这种隐写术并不是直接将需要传递的信息传送给对方，而是以“中国谜语”的形式隐藏在一些看似无关的字句里。当接收方收到这些“谜语”时，他需要将“谜语”猜解出来，才能知道原来的意思，这个过程其实就是秘密信息的提取过程。例如，在我国三国时杨修和曹操的故事就是这样：当时塞北进贡来一盒酥饼，曹操写了“一合酥”三字于盒上，放在台上。杨修看见了，竟和众人分食了。曹操心里恼怒，问“为何这样？”杨修答道：“你明明写了一人一口酥，我们不敢违背你的命令。”在这里，杨修就从“一合酥”三个字中提炼出了“一人一口酥”这样的隐藏信息。另外，在我国古代，除了文字，像图画、手势、动作等都可以隐藏信息，这些隐藏的信息只能被特定的人群所读取，而其他大多数人看到的只不过是再普通不过的一件东西罢了，这也是另一种意义上的信息隐藏。

上面叙述的这些技术，都是将秘密信息直接写在物质上，再利用不为人知的方法传

递出去。但是，若是有心人士从中作梗，例如将木制蜡板的蜡刮除，秘密信息是不是就马上被得知了呢？因此，除了信息隐藏的技术外，学者还提出另一个与信息隐藏技术相辅相成的方式——密码术。公元 1499 年时，Johannes Trithemius（1462—1516）撰写了一本名为《Steganographia》的密码专著，里面介绍了一个简单的密码术。首先选出一些无意义的文字，接着选出偶数位置的字符出来，再从选出来的字中取出偶数字符组成一个句子。例如，原始字符串：Parmesiel Oshurmi Delmuson Thafloin Peano Charustrea Melany Lyamunto；抽出的文字：OshurmiThafloinCharustreaLyamunto；取出偶数字符：sumTaliCauteLauto；隐藏的拉丁信息：Sum tali cautela ut。该例子透过文字的重组，巧妙地将秘密信息安排到无意义或有意义的文字数据中。值得一提的 Steganography 这个字，就是由 Trithemius 将 steganos 与 graphein 两个希腊字结合起来而成的。除了重组，字符打乱也是很常见的加密手法，例如，原始字符串为 An Apple a Day Keeps the Doctor Away，我们将奇偶字符分别抽出，得到奇数字符串：AApeDyepeteotrw；偶数字符串：nplaaKeshDcoAa。再将这两个字符串合成新字符串 AApeDyepeteotrwynplaaKeshDcoAa。当要取回原来的字符串时，只需将后半段字符串插入前半段字符串中间即可。

此外，大家都知道，人民币上的水印可用于鉴别真伪。纸张水印是随着手工造纸技术的发展而出现的。据文献记载，最早的纸张水印出现在 1292 年意大利的 Fariano，该城市在造纸工业中占有重要地位。在 13 世纪的末期，Fariano 大约有 40 个造纸厂分享纸张市场，他们制造不同规格、质量和价格的纸张。在造纸厂之间，技工之间以及经销商之间都存在激烈的竞争。当时任何一方都无法追踪产品的出处以及规格和质量标识。水印的引入消除了可能出现的混乱，并成为了一种很好的标识方法。在水印技术发明以后，它很快就传遍意大利，然后传到整个欧洲。尽管水印最初只是纸张商标或造纸厂的标识，但是很快它们就成为纸张格式、质量和强度的标识，并且也用来作为基本的日期和身份鉴别。一个证明水印法律效力的极好例子是 1887 年发生于法国的一个案例，水印被用作一项证据，证明字符已事先被嵌入，此案后来居然导致对代理人的起诉、警察长官解职、内阁倒台及总统辞职。

1.6.2 近代信息隐藏技术

人类进入 20 世纪后，又发明了很多方法用于信息隐藏：高分辨率缩微胶片、扩频通信、流星余迹散射通信和符号码等。另外，值得一提的是，想要了解密码术的历史，人们不可不读 1967 年出版的 David Kahn 的著作《破译者》（The Code Breakers）。在此之前，有关密码术的文献微乎其微。这本书使许多学者开始了解密码术，从而推动了密码术的研究。这本书中也记载了许多信息隐藏的例子，可供研究传统信息隐藏技术历史的人员阅读。

第二次世界大战时，德国人就发明了把胶片制作成句点大小的微粒技术，而放大后又能获得与标准大小页面相同的清晰度。德国军队就是使用这种微缩照片技术来将重要的秘密缩小成一个小点，再将小点伪装成标点符号放在信件中。虽然后来还是被美国联邦调查局所发现，但德军早已成功地将大量军事情报传递出去。严格地说，信息既没有隐藏也没有加密，仅仅是太小了而不被注意。用如此小的微粒胶片传送大量的数据（包括图画和照片等）的思想在网络时代仍值得借鉴。另外，在经典的间谍片中，人们经常看到用隐形墨水来隐藏信息。随着化学工业的发展，在第一次世界大战中人们制造出了复杂的化合物做成隐写墨水和显影剂。现代使用的化学隐形墨水基于的是不同化学物质

混合后所产生的化学反应。在 20 世纪的两次世界大战中德国间谍都使用过隐写墨水。

扩频通信和**流星余迹散射通信**多用于军事上,使敌手难以检测和干扰通信信号。根据 Shannon 在信息论研究中总结出的信道容量公式可知,增加信号带宽可以降低对信噪比的要求,当带宽增加到一定程度,允许信噪比进一步降低,有用信号功率接近噪声功率甚至淹没在噪声之下也是可能的。**扩频通信**(Spread Spectrum Communication)可看作是一种把信息隐藏在宽频伪随机噪声中的通信方式。它使用比发送的信息数据速率高得多的伪随机编码,扩展作为基带信号的信息数据频谱,成为极低功率谱密度的宽带信号,从而在实际上难以和背景噪声相区别。接收端使用相关处理方法,从收到的宽带扩频信号恢复基带信号。由于伪随机编码和信号相关处理这两大特点,使扩频通信具有较强的抗检测性,抗干扰性和保密性。一般而言,与宽带的载体信号相比,待隐藏的消息为窄带信号,用扩频技术对消息作处理,就可使载体和处理后的消息的频带相匹配。此外,高频有利于嵌入消息的不可见性,但却不利于鲁棒性,低频尽管有利于鲁棒性,但却会带来不可接受的可见性。扩频技术可通过将低频能量信号嵌入到每一个频段来解决这种矛盾。**流星余迹散射通信**是利用对流层散射现象的通信方式。从地面到十几公里高空的大气层称为对流层。在对流层中由于大气的湍流运动产生了具有各不相同的介电常数的湍流团,当无线电波照射到这些不均匀的湍流团时,就在每一个不均匀体上感应电流,成为二次辐射体,从而向各个方向发出该频率的二次辐射波,这就是散射现象。对流层散射通信就是利用这种现象而实现的超视距无线电通信。由于对流层散射现象在 200~8000 兆赫频段比较显著,所以对流层散射通信主要工作在这个频段内。20 世纪 50 年代初,美国提出了建立对流层散射通信系统的设想,并于 20 世纪 50 年代中建立了对流层散射通信电路。中国于 20 世纪 50 年代中期开始研究对流层散射传播问题,20 世纪 60 年代初研制模拟对流层散射设备,20 世纪 70 年代开始研制数字对流层散射设备,并陆续建站投入使用。

符号码是指用非文字的东西来表示文字消息的内容,例如把手表指针拧到不同位置可表示不同的含义,用图画、乐谱等都可以进行语义编码。举例来说,在第二次世界大战中,检查者截获了一船手表,由于担心手表的指针位置会拼出一个秘密消息,他们在检查过程中对指针的位置进行了调整。这种利用手表指针位置来传递秘密信息的技术就属于符号码。

1.6.3 现代数字信息隐藏技术

现代数字信息隐藏技术的历史可以追溯到 1954 年, Muza 公司的 Emil Hembrooke 获得一项名为“Identification of sound and like signals”的专利,专利中描述了向音乐中嵌入不可感知的信号来证明所有权的方法。1983 年 Simons 提出了具有代表性的“囚犯问题”,此问题引发了人们对信息隐藏技术的关注。“计算机网络是现代密码术的母亲,而 Internet 就是现代信息隐藏技术的母亲”。世界上第一台通用电子计算机诞生于 1946 年,20 世纪 70 年代计算机网络的兴起掀起现代密码术的研究热潮,并使密码术发展成为一门相对成熟的学科。随着 20 世纪 90 年代 Internet 的迅速发展,多媒体技术的逐渐成熟和电子商务的兴起,网上多媒体信息急剧增加。如果没有网络,信息技术决不会有如此迅速的发展,而网络的开放性和资源共享使得网络信息安全问题日益重要。这直接导致了现代信息隐藏技术研究热潮的到来。不同于传统的“隐写术”,现代信息隐藏技术是基于计算机系统的各种隐藏手段和方法,它利用感知系统的冗余和数字多媒体系统的数据统

计冗余,将秘密消息以一定的编码方式或者加密方式嵌入到公开的数字媒体中。它所使用的载体对象一般是网络上传输的数字信息,例如文本、图像、视频、音频、多媒体等。这些隐藏的实现是基于现代密码术和计算机技术为前提的。

二十多年来,国外在信息隐藏方面的理论研究和实用技术方面发展很快,由于融合了密码术、信息论、应用数学、数字图像处理、计算机安全、管理科学、心理学、经济学以及法学等相关学科的大量研究成果,隐写术的研究在短短的几年间就走过了相当于密码术几十年的发展历程。第一篇关于图像数字水印技术的文章发表于1990年,而国际上正式提出信息隐藏的研究是从1992年Kurak等提出将图像降质用于秘密交换图像开始的。1996年上半年,在Ross Anderson的推动和组织下,在剑桥大学召开了国际上第一届信息隐藏研讨会,这次会议推动了信息隐藏的理论和技术研究。其后,分别于1998年在美国、1999年在德国、2001年在美国、2002年在荷兰、2004年在加拿大、2005年在西班牙、2006年在美国、2007年在法国、2008年在美国、2009年在德国、2010年在加拿大、2011年在捷克、2012年在美国在分别召开了第2~14届国际信息隐藏学术会议。到了2013年,该会议和另一个ACM重大国际会议—媒体和安全(ACM Workshop on Multimedia and Security)合并,称为第一届ACM信息隐藏和媒体安全国际会议(IH&MMSEC),已经在法国顺利召开,第二届于2014年在奥地利举行。针对数字水印技术这个研究分支,国际上也有一个专门的研讨会,即International Workshop on Digital Watermarking(IWDW),从2002年开始召开,至2013年一共已召开12届,从2011年开始改为数字取证和水印国际研讨会(International Workshop on Digital-forensic and Watermarking),缩写还是IWDW。此外,一些知名的学术组织,包括IEEE、ACM、SPIE和EURASIP等在它们主办的学术会议中设置专题或以杂志专辑形式对信息隐藏技术进行讨论。如Proceeding of IEEE于1999年7月出版了关于多媒体信息隐藏的专辑。

国内对信息隐藏的研究起步相对较晚,1999年12月,在何德全、周仲义和蔡吉人三位院士的积极倡导下,由北京电子技术应用研究所组织,召开了全国第一届信息隐藏学术研讨会。随后,2000年在北京、2001年在西安、2002年在大连、2004年在广州、2006年在哈尔滨、2007年在南京、2009年在长沙、2010年在成都、2012年在北京、2013年在西安又相继召开了第2~11届全国信息隐藏学术研讨会。这十一届信息隐藏会议有力推动了国内信息隐藏领域的发展。此外,2000年1月15—16日,国家863计划智能计算机专家组、中科院自动化研究所和北京邮电大学信息安全中心还成功地举办了数字水印技术研讨会。近10年来,国家863计划、973计划、国家自然科学基金等都对信息隐藏给予了资助。

目前,国外研究信息隐藏技术的学术机构有剑桥大学、麻省理工学院的媒体实验室、IBM研究中心等一些大学和机构,研究的重点在于如何将信息隐藏到图像、声音和文字等载体中。目前,对于信息隐藏技术应用在数字产品的著作权保护上(即数字水印技术)的研究较多。瑞士洛桑联邦工技院信号处理实验室和通信研究所、美国的NEC研究所、中国台湾国立交通大学等都作出了不少成就。除了学术界的研究之外,目前也有一些公司开发出一些软件如:Fraunhofer's SYSCOP、HIGHWATER FBI、Digimarc Corporation、DICE'S Argent Digital Watermark等,提供有关数字产品著作权保护的服务。国内研究信息隐藏的主要科研院所有:北京邮电大学信息安全中心,中国科学院自动化研究所模式识别国家重点实验室、哈尔滨工业大学、北方工业大学、清华大学、北京理工大学、北京电子技术应用研究所、国家信息安全测评认证中心、中山大学、浙江

大学、北京交通大学等单位。

随着学术研讨会的召开和信息隐藏研究的深入,信息隐藏技术及应用取得了长足的进展。但总体来说,信息隐藏技术尚未发展到完善得可实用的阶段,仍有不少技术性问题需要解决。信息隐藏技术发展到今天,还没有找到自己的理论依据,没有形成理论体系。信息隐藏技术要建立自己的理论体系,最为流行的做法是借鉴信息论和密码术中的有关概念和方法。国外的很多专家学者用信息论对信息隐藏技术给出了几种安全性模型,都试图想建立安全性的理论框架,从而对其进行深入的研究。但到目前为止,这些模型都没能得到其他学者的一致认可。目前,加密仍是网络上主要的信息安全传输手段,信息隐藏技术在理论研究、技术成熟度和使用性方面都无法与之相比,但它的潜在价值是无法估量的,特别是在迫切需要解决的版权保护方面,相信其必将在未来的信息安全体系中发挥重要作用。

1.7 信息隐藏的应用领域

信息隐藏的研究涉及多媒体处理、通信、密码术、数学、计算机科学和模式识别等多个领域,是一门新兴的交叉学科。信息隐藏技术的应用领域主要包括以下几个方面:保密通信、版权保护、版权跟踪、内容认证、标注和匿名通讯等。分别对其概括介绍如下。

1.7.1 保密通信

在网络全球化和经济全球化的现代社会中,每天都有大量的数据信息通过网络传送,其中会有涉及政治、军事、商业、金融和个人隐私等重要信息,一旦这些信息被非法截获将导致不可估量的后果。信息隐藏技术具有信息保密的作用,能很好地保护这些信息。保密通信主要用于秘密信息的安全通信,它所保护的是嵌入到载体对象中的秘密信息,通常把秘密信息隐藏在普通的多媒体信息中进行传输。由于网上存在数量巨大的多媒体信息,从而使得秘密信息难以被窃听者检测。美国 911 恐怖攻击事件后,美国国家安全部门在深入调查后发现,以本拉登为首的恐怖组织,很有可能利用信息隐藏的技术,将攻击行动、计划藏匿在网络的布告栏、讨论区、图片或新闻影片中。如果将隐藏技术与密码技术相结合,则可以同时实现数据的秘密传送和安全保护。这是因为:首先,隐藏在多媒体信息冗余空间中的数据具有不可感知性,从而保密通信的过程不易为窃听者所得知;其次,数据嵌入算法中带有密码作为控制参数,因此即使嵌入算法公开,其安全性仍能由密码来保证;再者,在嵌入之前一般要对秘密信息作加密处理,当然这个过程也是由密码所控制。

尽管信息隐藏技术起源于保密通信,但近几年来,由于互联网市场的迫切需求,数字多媒体水印技术和数字指纹技术已成为信息隐藏技术研究的重点。数字水印技术和数字指纹技术方面的研究是目前最为活跃的领域,主要用于版权保护、版权跟踪及真伪鉴别等目的,所要保护的是载体对象。与保密通信应用相比较,数字水印技术应用更加强调算法的鲁棒性。目前数字水印技术的应用主要包括如下五个方面:**版权保护、复制控制(Copy Control)、叛逆者/盗版者追踪(Traitor Tracing)、真伪鉴别、完整性鉴别**,分别在 1.7.2、1.7.3 和 1.7.4 节中介绍。

1.7.2 版权保护和复制控制

随着互联网和电子商务的迅猛发展,互联网上的多媒体信息急剧膨胀,数字化多媒

体产品也可通过下载的方式从网上直接购买，网络上提供的数字服务也越来越多，如数字图书馆、数字电视和数字报纸等。因为音频、视频和其他一些作品可以数字化形式出现，完美的复制品易于得到，导致大量未授权复制品的产生，从而受到音乐、电影、图书和软件出版业的广泛关注。由于数字化产品的版权容易受到侵犯，因此如何有效地保护这些数字产品的版权就成为一个极其关键的问题。**鲁棒水印技术**（Robust Watermarking）就能有效地解决这一问题，通过向数字媒体作品中嵌入隐藏的版权信息，可以在发生纠纷时为作者提供版权证明的依据。例如，在网上传播数字产品时，提供商向用户发送的是隐藏有双方信息代码的产品，其中含有的水印信息不能够被破坏。如果数字作品被非法倒卖，版权拥有者可以通过提取水印来证明版权。这里对数字水印技术的要求是必须对常见数据处理和攻击具有很高的鲁棒性。此外，还需要考虑其他一些要求，如：水印必须明确无歧义，并且在其他人嵌入另外的水印以后，仍然能够判断出正确的所有权。

要有效地保护版权，还需要有有效的技术手段，以使非授权者不能对数字产品进行非法复制。例如，DVD 协会征求关于版权标志的建议，用于加强对连续复制的管理。一种可行的基本思路是消费者可以使用 DVD 播放器不受限制地复制家庭录像和某些电视节目，但却不可以滥用于商业目的。在这个思路下，家庭录像应该是不做标志的，电视广播应该标志为“仅可复制一次”，商业影像标志为“不可复制”，兼容的消费设备就会遵照这些标志来自动操作。一种可行的技术方法就是在数字产品中嵌入反映复制状态的水印。例如，可把这种信息以嵌入水印的形式包含在 DVD 数据中。具有防复制功能的 DVD 播放器不允许回放或复制含有类似“禁止复制”水印信息的数据。对含有“允许复制一次”水印的数据只能复制一次，而不能多次复制。日本电气公司、日立制作所、先锋、索尼和美国商用机器公司等拟联合开发统一标准的基于数字水印技术的 DVD 影碟防盗版技术。新的防盗版技术在构成动态图像的每一个静态画面数据中，嵌入可防止数据复制的**数字水印**。这样，消费者可在自用的范围内复制和欣赏高质量动态图像节目，但以赢利为目的大批量非法复制则无法进行。

1.7.3 数字指纹（盗版者/叛逆者追踪）

上面提到的鲁棒数字水印技术通过在载体作品中隐藏版权信息和复制标志达到版权保护和复制控制的目的。鲁棒数字水印技术其实还有另外一个应用，即用于监视或追踪数字产品的非法传播和倒卖，这种应用通常称作**数字指纹技术**。与版权保护应用不同，数字指纹技术通过隐藏产品序列号来识别购买者，即数字水印技术用于提供版权证据来起诉盗版者，而数字指纹技术用于找到盗版者（即盗版者追踪）。**数字指纹**类似于软件产品的序列号，即在每个发行的产品复制中嵌入不同的水印，当产品被不当使用时，则可根据产品内的水印，查出违法的顾客是谁。因为单个加入水印的复制会受到**共谋攻击**（Collusion Attack），嵌入的水印必须被设计成共谋安全的。此外，在一些应用场合，如在 WWW 上用特定的网络搜索器搜索盗版图像，数字指纹的提取必须要简单、快捷。

1.7.4 内容认证（真伪鉴别、完整性鉴别）

内容认证的目的是检测对数据的修改，有时也叫真伪鉴别或完整性鉴别。通常，可以使用**脆弱水印技术**（Fragile Watermarking）来实现内容认证。为便于检测，**脆弱水印**（Fragile Watermark）对某些变换如压缩具有较低的鲁棒性，而对其他变换的鲁棒性更低。因而在所有的数字水印技术应用中，认证水印具有最低级别的鲁棒性要求。为了确

保作品的完整性, 侦测作品是否被他人篡改, 可以在作品中藏入验证用的信号, 当完整性受到质疑时, 则将信号取出用以验证作品是否有被修改, 或者标示被篡改的区域, 甚至可以利用该信号做进一步的作品修复。例如, 作者可以在著作中加入个人的签章或特殊符号, 当有仿冒品或需要辨识著作的真假时, 即可从著作中取出签章, 若无法取出原来藏入其中的信息时, 则表示该著作是仿冒品。

近几年来, 各种各样的网站如雨后春笋般不断涌现, 随之而来的网页内容的篡改和非法盗用问题也日益突出。在网页中加入合适的水印也许将成为保护网页、防止非法篡改和盗用的一种有效手段。另外, 随着电子商务的兴起, 在商务活动中会牵扯到数字票据, 但数字票据很容易被伪造。在数字票据中嵌入水印可以用来防伪。显然, 电子商务中各种电子票据的有效防伪是十分重要的。电子票据水印技术将在今后几年得到更多研究。此外, 水印技术用于印刷品的防伪在 17 世纪就已出现, 但将数字水印技术用于印刷品防伪则是近些年刚刚提出的新方案。该方案在数字图像印刷、打印之前先嵌入一定的秘密信息, 经印刷 / 打印输出后的纸张可以再次扫描输入, 利用特定的水印提取和鉴别算法来验证该图像作品的真伪或所有权。

1.7.5 标注

标注 (Annotation) 也叫注释或数据附加, 它在数字作品中嵌入一些附加信息, 这些信息可以是关于作品的细节、注释等。这种隐式注释不需要额外的带宽, 且不易丢失。此项技术常被用于医学图像维护, 当医护人员拍摄完病人的 X 光片或摄影片后, 可以将病人姓名、主治医生、病史等隐藏于图像数据中, 以避免图像数据与病人记录间的连接遗失时无法找到正确的病历记录。

1.7.6 其他应用

信息隐藏技术的其他典型应用如下。

1. 匿名通讯

有些国家或政府会限制国民在线交谈的权利, 或者管控民间使用的加密技术, 因而刺激了人们发展各项隐匿技术, 以进行匿名通讯。此外, 电子投票或电子货币也常会使用到匿名技术。互联网络的合法用户可能使用匿名通讯来使用求助热线或在在线选举中私下秘密地投票。在数字电话中, 采用临时的移动用户标志的目的是防止用户的位置被揭发。利用匿名转发邮箱可以隐藏电子邮件的发送者信息。例如有些商家使用电子邮件匿名技术, 既发送了大量不合法的消息, 又避开了气愤的用户反应。

2. 保证交易的不可抵赖性

目前, 网上交易日益增多, 交易双方的行为应当被明确。当发送者发送信息或者接收者接收信息时, 将各自的特征信息以鲁棒性水印的形式嵌入到传递的信息中, 可以达到确认交易行为的目的, 这样双方都将不能抵赖。

3. 数字签名

现代通信技术为人们的生活提供了相当的便利, 比如只需要用户的一个电话就可以实现银行转账等, 在这些情况下为了安全, 常常用户需要提供身份认证, 在识别技术尚不完善的今天, 数字签名技术可以为许多场合下的身份认证提供帮助。同样, 如果在数

字产品中嵌入数字签名信息，就可以通过从非法复制中提取出隐藏的信息确定信息的来源，为追查责任提供依据。

4. 防止发送者被敌方定位

军事机构或其他一些情报机构，通常需要“低调”的通讯手段。在现代军事通信中，一旦检测到信号，马上就可以对其进行攻击。这样一来，即使对消息的内容加密，现代战场上对这些敏感信号的检测可能导致对发报员的快速反击。正是由于这个原因，在军方通讯中，扩频调制和流星猝发通信等通讯安全技术越来越多地采用，使得信号很难被检测或者阻塞掉，保证信号不易于被敌方发现或者攻击。犯罪分子也关注和采用一些“隐蔽”的通讯手段。他们乐于使用预付费的移动电话、修改过身份的移动电话，甚至侵入交换机使得电话可以改变通讯线路。

1.8 本章小结

本章首先从网络信息安全的背景出发，引出信息隐藏的定义、术语和特性要求。然后介绍了信息隐藏的几种模型，包括囚徒模型、通用模型、通信模型、广义模型和基于信息空间的模型。接着，介绍了信息隐藏的各个主要研究分支，包括隐写术、数字水印技术、数字指纹技术、隐蔽信道、阈下信道、低截获概率通信和匿名通信。然后，对信息隐藏技术的分类作了全面介绍，主要讨论按照载体对象、密钥的对称性和嵌入域的分类方式。接着，介绍了信息隐藏的发展历史，分古代、近代和现代三阶段来阐述发展进程。最后，介绍了信息隐藏潜在的几个主要应用领域，包括保密通信、版权保护、复制控制、版权跟踪、内容认证和标注等。

总的来看，作为一个新兴的研究领域，信息隐藏应用的广泛性已经引起了国内外信息隐藏研究的热潮。但是，到目前为止，信息隐藏技术的理论基础还没有完整建立起来，在不同的应用背景和信道环境下，具体信息隐藏容量大小也还要在理论上做进一步探讨。在信息隐藏技术发展的同时，针对信息隐藏的攻击技术也在飞速发展。目前的信息隐藏系统中，尚没有一种能够在各种攻击下都表现出良好的健壮性因此仍需要对现有的隐藏算法的鲁棒性、安全性等特性进行研究，结合数字信号处理技术，提出更好的信息隐藏的切入点以及相关算法，并进一步提高信息隐藏的容量，使其在更加广阔的范围得到充分应用。



习题

1. 请简述信息隐藏的术语和一般通用模型。
2. 请指出信息隐藏的广义模型和不对称信息空间模型之间的区别和联系。
3. 请阐述一个好的信息隐藏技术必需满足哪些要求？
4. 请讲述几个历史上的信息隐藏实例。
5. 举例说明信息隐藏技术的典型应用。

6. 假设有一幅 256 灰阶的灰度图像的某 4×4 图像块如题图 1.1 所示。灰度图像是由许多像素 (Pixel) 所组合而成的，每一个像素值表示该点的灰度，256 灰阶指的是每个像素的取值范围从 0 到 255。如果像素是一个黑点，则灰度值为 0，如果像素是一个白

点，则灰度值为 255。因为灰度值的值域范围从 0 到 255，所以可用 8 位二进制数来表示一个灰度值，以黑点而言，其二进制表示为 $(00000000)_2$ ；若是白点则二进制表示为 $(11111111)_2$ 。若是灰阶值为 5 的像素，其二进制表示如题图 1.2 所示，其中最左边的位称为最重要位（Most Significant Bit, MSB），而最右边的位为最不重要位（Least Significant Bit, LSB）。由于最不重要位 LSB 对像素的贡献度最小，即改动该位对整个像素值的取值不会有太大影响，人类的视觉系统一般查觉不出这么细微的改动。因此，许多替换式隐藏技术，就是替换 LSB 部分将信息藏入图像中。假若要将 16 位机密信息 $H=1101011110101101$ 嵌入题图 1.1 所示的像素块中，请给出伪装图像块的像素值。

5	10	11	12
20	31	40	51
27	10	17	45
37	85	15	34

题图 1.1 某 4×4 灰度图像块

MSB						LSB	
2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
0	0	0	0	0	1	0	1

题图 1.2 像素值 5 的二进制表示

第 2 章

隐写术

本章引言

隐写术是信息隐藏领域中一个较早出现的重要分支，它是一门将秘密信息嵌入到看似平常的其他载体信息中进行传送以防止第三方检测出秘密信息的技术。隐写术（Steganography）和密码术（Cryptography）两个词的正式出现都是在 17 世纪中叶并且都是来源于希腊语，它们都主要用于保密通信。本章首先概述保密通信的有关背景，接着概述隐写术的有关概念、分类和性能评价问题，然后按照载体类型的不同分别介绍基于文本、图像、音频和视频等载体的隐写术。

本章重点

- 隐写术的相关概念和分类；
- 隐写系统的性能评价；
- 基于图像载体的隐写术；
- 基于音频载体的隐写术。



2.1 保密通信概述

2.1.1 基本概念和分类

保密通信是一种通信对象之间为防止秘密信息被窃取，按约定的方法改变信息的传输方式或表现形式以隐蔽其真实内容的通信方式。通常，保密通信系统对传输的秘密信息在发送端进行加密处理或隐藏处理，在接收端进行解密处理或提取操作来恢复原始秘密信息，使窃密者发现不了秘密信息的传输或者即使截获了传输的信号，也不了解信号所代表的信息内容或隐藏其中的信息内容。

保密通信技术一般可分为线路保密技术和信息保密技术两大类。线路保密技术研究的重点是如何不让敌方收到本方的通信信号。线路保密可以称为隐蔽式的保密通信，它使敌方不能从本方通信线路（包括有线和无线线路）上直接获取信号，或者说以不为敌方所觉察的方式来进行通信。例如无线通信中的定向发射、窄向发射等通信技术，就是属于这种方式。从线路保密性来看，以光纤通信为最好（在约 10km 长的光纤线路中，仅能在距离 4~5m 的范围内收到信号）。在有线通信的线路保密技术中也可以使用特制的保密电缆，但其价格比较昂贵、难以推广。实践证明，单靠线路保密技术，是很难防止敌对方窃密的，为了最大限度地确保通信的安全，必须结合使用信息保密技术。应该说现代通信的保密，主要还是依靠信息保密技术来实现的。信息保密技术研究的重点是不让敌方译出本方的通信信息，即研究如何对传送的信号进行变换、加密或隐藏。这样在通信过程中，敌方虽可通过各种方式窃收到本方信号，但却解不出信号的真实内容。

根据保密通信的技术手段，可以将信息保密技术分为基于经典密码术的保密通信、低截获概率通信、基于混沌理论的保密通信、基于量子通信的保密通信、基于隐写术的保密通信。在古代通信中，人们已采用暗号、隐语、密码等保密措施。中国古代兵书《六韬·龙韬》记载，西周时期曾用过以约定的物体长短表示某种意义的“阴符”方法，和将一份文书分解成三个部分分别传送的“阴书”方法。在古罗马帝国凯撒时期（公元前 59—前 44）就曾采用密码方法。由于电子技术的发展，现代通信中已广泛使用密码术和隐写术进行保密通信。低截获概率通信是使信号被截获的概率降低的通信技术，主要包括扩展频谱通信技术和流星猝发通信技术。混沌保密通信是利用混沌信号所具有遍历性、非周期、连续宽带频谱、似噪声等特性而进行的保密通信。量子密码术是量子物理学和密码术相结合的新兴交叉科学，它成功地解决了传统密码术中单靠数学无法解决的问题并引起国际上的高度重视。不同于基于经典密码术的通信，这种新的保密通信基于量子效应，有着绝对安全的优点。隐写术以秘密的方式传递明文，通常将秘密消息隐藏到其他载体中，目的是隐蔽消息的存在。为了增加保密性，通常对秘密消息先加密再隐藏。

保密通信技术还有其他分类方式。若按结构原理分，包括调制式保密通信、频分式保密通信、时分式保密通信、伪装式保密通信、组合式保密通信和加密式保密通信等。若按通信方式分，包括电报保密通信、传真保密通信、电话保密通信和电视保密通信等。若按传输线路分，包括有线保密通信和无线保密通信。下面概要介绍基于经典密码术的保密通信、混沌保密通信、量子保密通信和基于隐写术的保密通信。

2.1.2 基于经典密码术的保密通信

基于经典密码术的保密通信是指利用对称或非对称密码体系进行的保密通信。在对

称加密算法中, 通信双方共享一个不对外公开的密钥 (即私钥), 依靠同一个密钥进行加密和解密数据, 最典型的对称加密算法是 DES。虽说窃听者可以通过穷举方式来试探密钥, 但由于其计算量将随密钥长度的线性增加而呈指数级增加, 故我们可以通过增加密钥的长度来提高其安全等级。例如, 用一个 128bit 长的密钥对明文加密, 如果采用野蛮攻击方法, 需要 10 亿台运算能力为 10 亿次/秒的计算机花费 10 000 亿年的时间才能解密。在对称加密算法中, 密钥必须很好地保管, 才能够保证密文的安全性。如果攻击者得知了用户的密钥, 他就能解密所有的加密信息。因此, 需要一种安全的方法对密钥进行保密传输, 这就是通常所说的“密钥分发”问题。

为了解决这个问题, 人们又提出了非对称加密的方案。所谓非对称加密 (又称公开密钥加密) 算法, 它使用一对密钥来分别完成加密和解密操作。一个密钥公开发布, 称为公钥; 另一个由用户自己秘密保存, 称为私钥。发送者用对方的公钥去加密, 而接收者则用自己的私钥去解密。人们通过数学的手段来保证加密过程是一个不可逆的过程, 即用公钥加密的信息只能用与该公钥配对的私钥才能解密, 最典型的算法是 RSA (Rivest、Shamir 和 Adleman 三人提出来的一种非对称加密算法)。

尽管在量子计算机时代来临时, 以 DES 为代表的对称加密算法和以 RSA 为代表的非对称加密算法面临着安全性全部失效的威胁, 一次一密的方案仍是一种有效的保密方式。只要私钥没有被窃取, 密文就不会被解密。在一次一密方案中, 随机生成一个与明文长度相等的密钥, 再将两者简单相加便可得到密文, 这样所得的结果是完全随机的, 窃听者不可能从所得的结果中得到任何有用信息。不过要强调的是, 这种密钥是不可以重复使用的。因为窃听者可以通过分析两段使用同一密钥加密的密文来分解出密文, 这样密文就无法保证是安全的了。理论上, 现在只有一次一密的方案是一种无条件安全方案, 其安全性不依赖于破解者的计算能力, 无论密码分析者拥有多么巨大的计算能力, 在没有密码本的情况下, 都不能把信息恢复出来。但是, 一次一密方案要求密钥长度与明文长度相等, 因此会增加通信带宽和存储负担, 然而更严重的是密钥分发问题。虽然在理论上要求接收方和发送方共享的密钥是保密的, 但实际中是不可能做到的。密码本可能在传输过程中泄密, 因为窃听者在经典的物理信道上监听时, 不会对信道的物理特性产生任何影响, 他可以在收发双方毫不知情的情况下, 任意篡改信道上传输的信息。毫无疑问, 这种情况下保密通信是无安全性可言的。

2.1.3 混沌保密通信

混沌保密通信是 20 世纪 90 年代初伴随着混沌同步现象的发现而发展起来的新型混沌应用科学。如今它已经成为保密通信领域中的一个重要研究方向, 主要因为混沌信号具有如下几个重要特点。首先, 混沌信号具有非周期、宽频谱、似噪声的特性, 与纯粹的随机过程毫无区别。窃听者无法利用频谱信息来对混沌信号进行跟踪分析, 往往误认为噪声而加以忽略, 而其独特的初始条件的敏感依赖性, 决定了混沌信号的长期行为不可预测, 从而达到保密通信的目的; 其次, 混沌序列及其变换序列具有更大的复杂度。复杂度即序列的等效线性长度, 根据密码学原理, 复杂度越大, 系统就越难破译, 根据混沌信号的遍历性, 混沌信号可以遍历有界混沌区域内每一个状态而绝不重复, 决定了其产生的混沌序列具有无限长度, 解决了伪随机信号有限周期的局限。另外, 混沌信号种类繁多、数目巨大。混沌信号是由确定的非线性系统产生的, 由于混沌系统是自然界普遍存在的一类现象, 所以不用担心混沌系统会使用殆尽。新的混沌系统不断被发现保

证了混沌系统这一丰富的信息源能够在保密通信中获得广泛的应用。因此，混沌通信及相应的信息加密技术很具竞争性，已成为 21 世纪大有发展前景的高新科技研究领域。

2.1.4 量子保密通信

量子通信是目前科学界公认的唯一能实现绝对安全的通信方式，因为它能及时检测到窃听行为，从而为通信双方提供一个无条件安全的通信系统。量子通信的保密原理与传统的保密原理有着根本的不同，量子保密通信的理论基础是量子力学，而以往密码术的理论基础是数学。量子密码术利用物理学原理保护信息。首先想到将量子物理用于密码技术的是美国科学家威斯纳，威斯纳在“海森堡测不准原理”和“单量子不可复制定理”的基础上，逐渐建立了量子密码的概念。“海森堡测不准原理”是量子力学的基本原理，指在同一时刻以相同精度测定量子的位置与动量是不可能的，只能精确测定两者之一。“单量子不可复制定理”是“海森堡测不准原理”的推论，它指在不知道量子状态的情况下复制单个量子是不可能的，因为要复制单个量子就只能先做测量，而测量这一量子系统就会对该系统产生干扰并且会产生出关于该系统测量前状态的不完整信息。因此，窃听量子通信信道就会产生不可避免的干扰，合法的通信双方则可由此而察觉到有人在窃听。量子保密通信利用这一原理，使从未见过面且事先没有共享秘密信息的通信双方建立通信密钥，然后再采用 Shannon 已证明的是完善保密的一次一密钥密码通信，即可确保双方的秘密不泄漏。

2.1.5 基于隐写术的保密通信

基于隐写术的保密通信是本章介绍的重点。它是将秘密信息隐藏到载体对象中，以产生伪装载体，因为伪装载体和原始载体非常相似，所以伪装载体在传输过程中，不会引起其他人的怀疑。下面各节首先介绍隐写术的概念、分类和特性要求，然后按照不同载体类型分别介绍常见的隐写术。

2.2 隐写术的相关概念和分类

隐写术和数字水印是信息隐藏中最为重要的两大分支，两者之间紧密相关又有重要区别。这两个领域重叠面较大，共享很多技术方法，但在基本哲学涵义上存在着差异，并影响到技术方案的要求与设计。本章介绍隐写术，第 3 章将介绍数字水印技术。

2.2.1 隐写术的基本概念

隐写术 (Steganography) 一词来源于希腊词汇 *stegnos* 和 *graphia*，意即“隐藏”(Cover) 和“书写”(Writing)，是一种保密通信技术。隐写术是一门古老的技术，它将秘密信息嵌入到看上去普通的信息中进行传送，以防止第三方检测出秘密信息。隐写术在其发展过程中逐渐形成了两大分支，分别为**语义隐写术** (Linguistic Steganography) 和**技术隐写术** (Technical Steganography)，后者是本章讨论的重点。

隐写术的一般原理如图 2.1 所示。Alice 想把**秘密信息 (嵌入对象)** m 传送给 Bob，她首先随机地选取一个**载体对象** c (取自私有的随机信源，它可以不受怀疑地传送给 Bob)，然后在 c 中隐藏秘密信息 m ，可能还要使用**隐写密钥 (隐藏密钥)** k 。由此 Alice 将载体对象 c 改变成**隐写对象 (伪装对象)** s 。Alice 必须非常小心地进行这些工作，以

使得第三方 Wendy 仅能知道表面的载体信息而无法检测到秘密信息的存在。无论是对人还是对采用统计方式的计算机而言，一个“完美”隐写系统中的载体对象和隐写对象应当是不可区分的。理论上讲，载体对象是任何计算机可读的数据，如数字图像、数字音频或文本。Alice 在公开信道上将 s 传送给 Bob，希望 Wendy 不会注意到所嵌入的数据。因为 Bob 知道 Alice 所采用的嵌入方法和嵌入过程中所使用的密钥 k ，故他能提取 m ，且该提取过程在没有载体对象 c 的情况下也能进行。正在监视通信的第三方 Wendy 应当不能决定发送方送出的隐写对象中是否含有秘密信息，换句话说，即使第三方能掌握通信双方所传送的载体对象集合 $C=\{c_1, c_2, \dots\}$ ，他也应该不能决定哪一个 c_i 中含有嵌入信息，从而，隐写系统的安全性主要取决于第三方区分载体对象和隐写对象之间差别的能力。但是，在实际应用中，并非所有的数据文件都可以用作保密通信中的载体对象。这是因为要保证让未参与通信过程的人发现不了在数据嵌入过程中所做的修改，就要求载体对象含有足够多的冗余以便能被秘密信息所替换。很显然，同一载体对象不应当被使用两次，因为攻击方（第三方）如果得到同一个载体对象的两个不同“隐写版本”，就能很容易地检测到并且可能恢复出秘密信息。为了避免意外的重复使用，发送方和接收方都应当销毁已在信息传送中使用过的载体对象。

隐写术和密码术不同，密码术试图隐藏信息的内容，而隐写术则更进一步，它试图隐藏通信事件本身的存在性。在隐写术中，收发双方在隐蔽地交换秘密信息（包含加密形式的秘密信息）时，他们都得考虑被动的、主动的或者是恶意的攻击者。在一个隐写系统中，攻击者的主要目的在于正确检测出秘密信息的嵌入位置，对于更强有力的看守者来说，还可查明具体的秘密信息。如果看守者 Wendy 只能观察 Alice 和 Bob 之间的通信，则称之为**被动看守者**；反之，如果 Wendy 既可以观察又可以修改流经他的消息，则称之为**主动看守者**。主动看守者还可以向第三方证明 Alice 和 Bob 之间传递的隐写对象中存在秘密消息，甚至指出秘密消息的内容，也可以在不改变隐写对象的前提下，从隐写对象中删除秘密消息，但是看守者不可以对消息进行阻塞，即删除所有可能的秘密消息而不考虑载体对象，因为这样的话他就侵犯了 Alice 和 Bob 的人身权利。

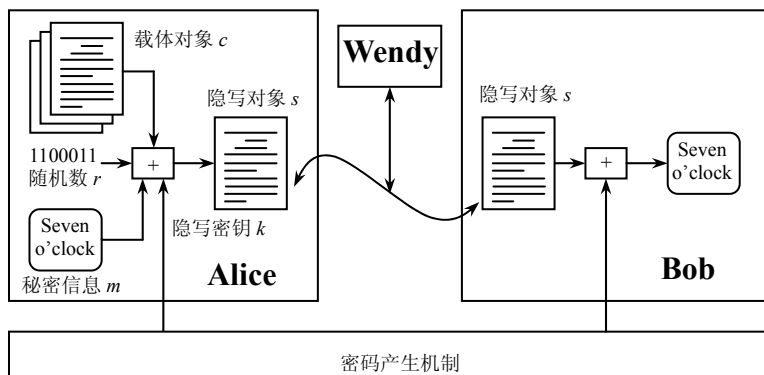


图 2.1 典型隐写系统模型

2.2.2 隐写术的分类

要达到隐蔽传送秘密信息的目的，可以使用各种各样的技术手段，根据其实现的主要思想也可将隐写术分为基于隐匿安全的隐写术、基于伪装的隐写术、基于隐藏被嵌入信息位置的隐写术、基于扩散被嵌入信息的隐写术，另外还有一些特殊环境下可以使用

的隐写技术。语义隐写术包括**符号码**（Semagram）、**隐语**（Open Code）以及**虚字密码**（Null Cipher）等。技术隐写术有多种分类方法，目前主流的分类方法有三类，分别按照隐写系统结构、隐写空间、载体对象类型进行分类，如图 2.2 所示。

2.2.3 语义隐写术概述

语义隐写术利用语言文字自身及修辞方面的知识和技巧，通过对原文进行一定规则的重新排列或剪裁，从而隐藏和提取密文。语义隐写术包括符号码、隐语以及虚字密码等。

1. 符号码

所谓**符号码**（Semagram）是指一次非书面形式的秘密通信。从实现思想上看，符号码属于基于隐匿安全的隐写术。举例来说，Schott（公元 1608—1666）在著作《Schola Steganographica》中阐述了如何在音乐乐谱中隐藏信息，每个音符对应于一个字符，如图 2.3 所示。Schott 还扩展了 Trithemius（公元 1462—1516）在《Steganographice》一书中提出的“Ave Maria”码。扩展码使用 40 个表，每个表有 24 个入口，其中每个入口对应于当时字母表中的一个字母，这些入口包括四种语言：拉丁文、德文、意大利文和法文。纯文本中的每个字母，被相应入口内的词或短语所替代，最终隐写文本看上去像是祈祷词或者咒语。在第二次世界大战中，曾利用图片中草的长叶片作为莫尔斯码的划线，短叶片作为圆点（图 2.4）来传递秘密信息，该方式也属于语义隐写术中的符号码。需要注意的是，语义隐写术要达到不引人注意的目的，在载体对象的选择上应该注意一定的技巧。举例来说，如果直接用音符来代表不同的字母，那么符号码产生的隐写乐谱就很可能看起来和听起来都根本不像音乐，这样的操作过程并不是一次成功的隐写。

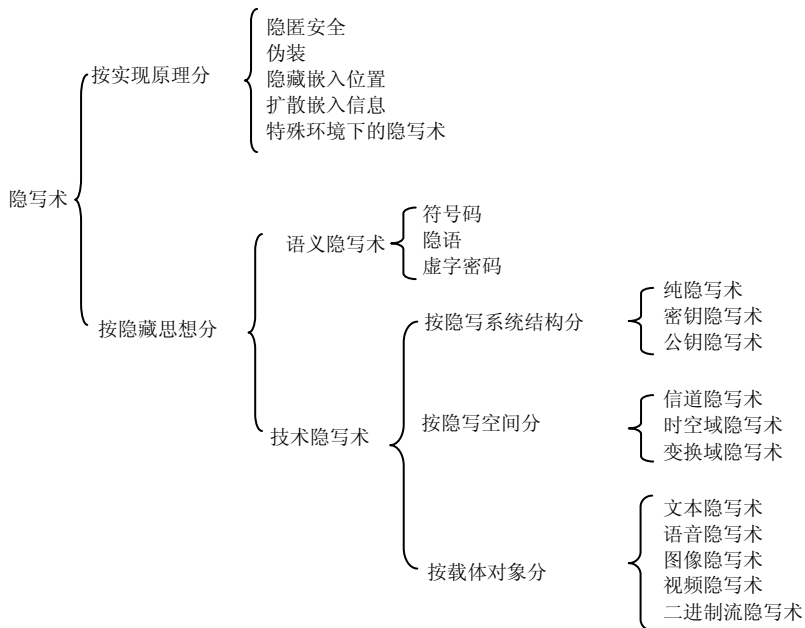
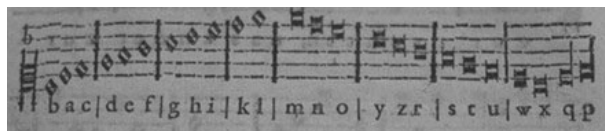


图 2.2 隐写术分类

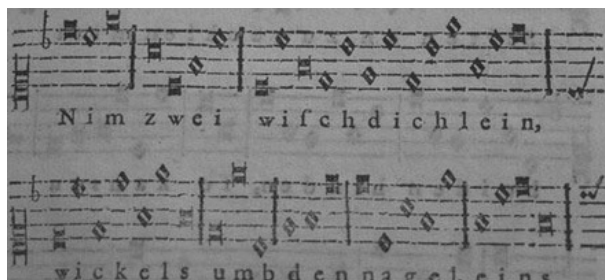
2. 隐语

隐语（Open Code）所利用的则是错觉或代码字。从实现思想上看，隐语属于基于隐匿安全的隐写术。例如，在第一次世界大战中，德国间谍使用雪茄的假订单来代表不同

类型的英国军舰——巡洋舰和驱逐舰，例如，朴次茅斯需要 5000 根雪茄就代表着朴次茅斯有 5 艘巡洋舰等。另外，在第二次世界大战期间，一个名叫 Valer Dickinson 的妇女使用玩偶作为代码字表示美国在纽约的船只数目来向日本发送信息，她使用小玩偶代表驱逐舰，而用大玩偶代表航空母舰或战列舰。

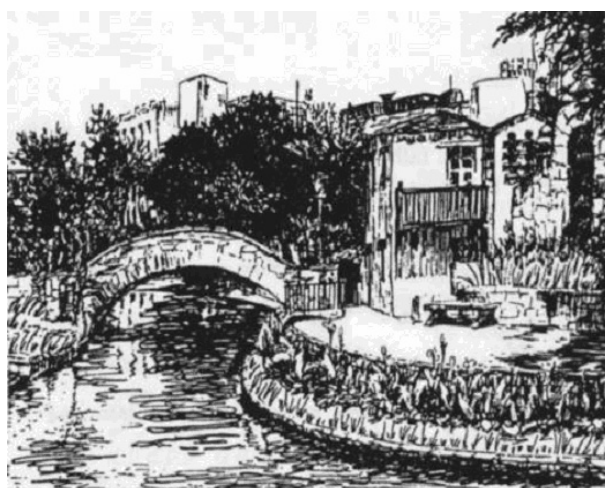


(a) 字母到音符的映射



(b) 隐藏有信息的乐谱

图 2.3 在音乐谱中隐藏信息



(a) 整幅画



(b) 沿河岸的草叶（短草叶代表莫尔斯电码的点，长草叶代表划）

图 2.4 隐藏着一封密信的圣安东尼奥河面画

3. 虚字密码

在**虚字密码**（Null Cipher）中，使用每个单词的相同位置的字母来拼出一条消息，但是这样的载体文本非常难以构造并且听起来会比较奇怪。如果构造者有足够的时间和空间，则可以通过精心设计来减少一些奇异性。最著名的以隐藏形式表示的虚字密码的实例是 1499 年由佚名人士写的令人费解的《Hypnerotomachia Poliphili》一书。该书揭示

了一个修道士和一个女人之间的罪恶的爱情，将其 38 章标题的第一个字母拼起来可得一条信息“Poliam frater Franciscus Columnaperamavlt”。在 Kahn 的《The Codebreakers》一书中，他列举了一个修道士是如何写下一本书，并把他的心上人的名字设计为连续章节标题的第一个字母拼接结果。另外，我国古代经常使用的“藏头诗”也是虚字密码的一种形式。在一首诗中，将各行的首个字连接起来代表一条秘密消息。

4. 几何隐语

几何隐语 (Geometric Open Code) 最古老的形式是万向网格。从实现思想上看，几何隐语属于基于隐藏嵌入位置的隐写术，即在一张空白纸上随机剪出与秘密信息单词长度相当的孔洞，并制作该张纸的副本发给接收方。编码者将他的网格放在一张空白纸上，然后在孔洞中写入秘密消息，随后拿开网格，在孔洞之间的空白处用其他单词填满，并将信件发送出去，接收者将他的网格放置在信上就可以读出秘密消息。但是，与虚字密码相同，在几何隐语中构造一个读起来不引人注意的载体消息也是比较困难的，没有经过精心设计的网格消息总是听起来很古怪，因而也常常容易被检测者所发现。

2.2.4 技术隐写术概述

技术隐写术是隐写术中的主要分支，历史上最著名的一个技术隐写术实例是发生在大约公元前 440 年被称为“剃头刺字”的故事。毫无疑问，技术隐写术的发展是伴随着科技，尤其是信息科技的发展而发展的。从古代利用动物（如兔子、狗）的身体及在木片上打蜡，到近代使用的隐形墨水、缩微胶片，再到当代使用的扩频通信、网络多媒体数据隐写等，可以说每一种新隐写术的出现都离不开科技的进步。

在过去 20 年中，人们已提出了许多不同的现代信息隐写术，其中许多技术都是基于替换方法，即用秘密信息替换载体信息的冗余部分，其主要缺点是隐写算法的鲁棒性相对较差。近年来，鲁棒数字水印技术的发展带动了鲁棒和安全信息隐写系统的发展。下面分别按照不同的分类方法介绍各种技术隐写术的基本原理。

1. 纯隐写术、密钥隐写术和公钥隐写术

(1) 纯隐写术 (Pure Steganography)

纯隐写术就是不需要预先交换附加信息（如隐写密钥）的隐写术。纯隐写术可定义为四元体 $\{C, M, E_m, E_x\}$ ，其中 C 是所有载体对象的集合， M 是有可能信息的集合，满足 $|C| \geq |M|$ ，即载体对象数目应该大于所有可能消息的数目。注意，这里 C 也包含所有可能的隐写对象。这样，嵌入过程 E_m 可用映射 $C \times M \rightarrow C$ 来描述，提取过程 E_x 可用映射 $C \rightarrow M$ 描述，且满足对所有的 $m \in M$ 和 $c \in C$ 有 $E_x(E_m(c, m)) = m$ 。收发双方都必须掌握嵌入和提取算法，且算法不公开。在纯隐写术中，除函数 E_x 和 E_m 之外不需要其他信息即可启动通信过程，系统的安全性完全取决于隐写过程本身的安全性。

(2) 私钥隐写术 (Symmetric Steganography)

显然，纯隐写术违反了 Kerckhoff 原理（即隐写术和密码术的安全性必须依赖于密钥的安全性，算法必须公开），因而并不安全。私钥隐写系统和对称加密系统相类似，发送方选择一个载体对象 c 并利用密钥 k 将秘密信息嵌入到 c 中。如果接收方知道嵌入过程中所使用的密钥，则他能进行逆过程以提取信息，其他任何不知密钥的人都不能获取秘密信息。**私钥隐写术**可定义成五元体： $\{C, M, K, E_x, E_m\}$ ，其中 C 为载体对象集合， M 为秘密信息集合，且满足 $|C| \geq |M|$ ， K 是密钥集合。 $E_m: C \times M \times K \rightarrow C$ 和 $E_x: C \times K \rightarrow$

M 具有下列性质: 对所有的 $m \in M$, $c \in C$ 和 $k \in K$ 有 $E_x(E_m(c, m, k), k) = m$ 。与密码术一样, 密钥隐写术也存在密钥交换问题, 通常假设通信各方能通过安全信道传送密钥。

(3) 公钥隐写术 (Asymmetric Steganography)

公钥隐写术不依赖于密钥的交换。公钥隐写系统需要用到两种密钥: 私钥和公钥, 其中公钥存放在公钥数据库中。公钥用于嵌入过程, 而私钥则用于提取秘密信息。假定 Alice 和 Bob 在入狱前就交换了某一公钥加密算法的公开密钥, 且加密算法和嵌入函数都是公开的。Alice 用 Bob 的公钥对信息加密, 得到外表随机的信息, 然后嵌入到 Bob 知道的载体中 (当然 Wendy 也知道), Bob 用私钥对收件进行提取和解密。即使 Wendy 能提取出 Alice 发送给 Bob 的密文信息, 但是由于加密过程中所产生的密文看上去是随机的, 使得 Wendy 没有理由怀疑它不是随机数据, 除非她想到去攻破而且能够攻破该密码系统。但是, 一个严重问题是 Bob 必须对 Alice 发来的每一份载体对象进行解码, 甚至他可能并不认识 Alice。如果嵌入对象不是专发给某人的, 例如是 Internet 的新闻节目中邮寄的, 则问题会更糟糕。

可以设想, 当有恶意的中介人存在时, 无论公钥隐写术或纯隐写术都行不通。Wendy 可能以 Alice 的名义启动一个公钥隐写方案来欺骗 Bob, 这种情况和需要验证公钥的公钥加密术是一样的。在纯隐写术中, Bob 无法区分信息是来自 Alice 还是 Wendy。

2. 时空域隐写术和变换域隐写术

(1) 时空域隐写术

时空域隐写术多采用替换法, 即用秘密消息位替换载体对象中的最不重要位。接收方只要知道秘密信息嵌入的位置就能提取信息。由于在嵌入过程中只作了很小的修改, 发送方可假定**被动攻击者** (Passive Attacker) 是无法觉察到的。

(2) 变换域隐写术

时空域修正技术是较容易实现的信息嵌入方法, 但攻击者使用简单的信号处理技术就可能完全破坏秘密信息。在许多场合, 有损压缩带来的小改变就可能导致整个信息的丢失。在现代隐写术发展早期, 人们就注意到把信号嵌入到频域中比嵌入到时域中鲁棒性更强。现有的多数鲁棒隐写系统都是在某个变换域中进行的, **变换域隐写术**把消息隐藏在载体对象的变换域系数中。与时空域方法相比, 变换域方法对诸如压缩、修剪等攻击的鲁棒性更强, 当然, 它们仍然是人类感官系统无法觉察到的。

3. 信道隐写术

利用信道的一些固有特性进行信息隐写的方法称为**信道隐写术**。目前发展起来的信道隐写术主要有以下两大类: 基于网络模型的信道隐写术和扩频隐写术。

(1) 基于网络模型的信道隐写术

网络中易失数据的信息隐写指利用网络中报文的控制数据和时序特性来隐藏信息数据, 即秘密信道可存在于网络的模型结构中。其特点: ① 数据是易失的。与在数据文件 (如数字图像、数字音频) 中进行隐写不同, 它们分别利用网络的控制信号或通信协议等媒介, 在通信进行过程中“夹带”了保密通信。一旦通信结束, 秘密信息随之消失, 因而不知情的第三方无法发觉保密通信的存在。② 从保密方法来看, 利用网络的控制信号或通信协议等媒介中的一些固定空闲位置或信号进行秘密信息传送, 要使这种保密通信不为他人所发觉, 须使出现在那些位置的数据呈良好的随机或伪随机特性。因此, 在发送前对秘密信息进行 (伪) 随机调制不失为一种好办法, 以增加通信保密性。

(2) 扩频通信

扩频通信 (Spread Spectrum Communication) 技术诞生于 20 世纪 50 年代, 它为我们

提供了一种低检测概率、抗干扰的通信手段。Pickholtz 等人把扩频技术定义为“一种传输手段，信号的带宽超过发送信息所需的最小要求。宽频是通过与数据无关的编码实现的，接收方同步地接收该码以用于解扩及随后的数据恢复”。尽管传输的信号功率很高，但每一个频段的信噪比较低。即使几个频带的部分信号可能被去除，其他频带中仍有足够的信息用以恢复信号。因而，扩频通信技术使得检测和去除信号变得困难。这非常类似于信息隐写系统，后者把秘密消息扩散到载体对象中以使之无法被感知。由于扩频信号难以去除，基于扩频通信技术的嵌入方法应该有较强的鲁棒性。自从 1993 年 Tirkel 等人的划时代文章“Electronic Watermark”发表以来，扩频方法在信息隐藏领域的重要性不断增加。信息隐藏中通常使用的两种扩频方法为直接序列方法和跳频方法。在直接序列方法中，秘密消息由一个称为片率（Chip Rate）的常值扩散，与一个伪随机信号调制，再加入载体对象中。而在跳频方法中，载波信号的频率快速从一个频率跳到另一个频率。

2.3 隐写系统的性能评价

一个现代隐写系统的性能评价主要包括透明性、秘密信息的正确恢复率、隐写容量、系统复杂度和安全性五个方面。下面对这五个方面分别进行阐述。

2.3.1 透明性

在绝大多数应用中，透明性是信息隐藏系统的基本要求，尤其对于隐写系统和数字水印系统。透明性也叫不可感知性，它是指嵌入秘密信息 $m \in M$ 不能使载体对象 c 产生可感知的失真，即隐写对象 s 和载体对象 c 应充分接近， $s \in C$ ， $c \in C$ 。对于一个隐写系统来说，具备了透明性才能使得隐写对象不会被怀疑。例如，若载体对象 c 为一幅图像，则肉眼应无法区分隐写图像 s 与载体图像 c 之间的差异；若载体对象 c 为一段声音，则人耳应听不出隐写声音 s 和载体声音 c 之间的区别，否则就失去了隐藏的意义。

1. 以音频载体为例

对于音频载体，透明性可以用隐写对象的听觉质量来刻画^[11]。该听觉质量指的是在嵌入了秘密信息后，人耳对隐写对象中的秘密信息的察觉度。该方面主要采用主观质量标准，特别是 MOS（Mean Opinion Score）得分来衡量，即人耳是否能察觉到隐写音频与载体音频相比有明显的听觉降质。MOS 得分方法是由 CCITT（Consultative Committee for International Telephony and Telegraphy）推荐的主观方法，现已广泛作为不同系统之间的比较标准。它采用 5 级评分标准，见表 2.1。MOS 得分的测试方法为：参加测试的实验者在听完所测话音后，从这 5 个等级中选择其中某一级作为他对所测话音质量的评定，全体实验者的平均分就是所测话音质量的 MOS 分。

表 2.1 MOS 得分说明表

MOS 得分	话音质量	失真程度
5	优	不易察觉
4	良	刚刚察觉但不讨厌
3	一般	可察觉但稍微讨厌
2	差	讨厌但能忍受
1	极差	非常讨厌且不能忍受

在实际应用中,也常用 ITU-TE862 来近似主观 MOS 得分。PESQ 是针对语音质量的直观评价标准,取值范围为 1.0 (最差)~4.5 (最好)。PEAQ 则用于估计音频文件的听觉降质情况,其输出值 ODG (Objective Difference Grade) 范围为-4 (极其明显的降质)到 0 (降质无法察觉)。

隐写音频的分段信噪比 (Segmental Signal-to-Noise Ratio, SNR_{seg}) 是另一个常用的客观评价方法,评判藏有秘密信息的隐写音频相对于原始音频的波形失真程度

$$\text{SNR}_{\text{seg}} = \frac{1}{N} \sum_{i=1}^N \text{SNR}_i \quad (2.1)$$

其中 $\text{SNR}_i = 10 \lg \frac{\sum_{j=1}^P c_i^2(j)}{\sum_{j=1}^P [s_i(j) - c_i(j)]^2}$ 为第 i 个音频数据帧 c_i 的信噪比,每帧 P 个样点,共 N 帧。 $c_i(j)$ 和 $s_i(j)$ 分别表示第 i 帧载体音频的第 j 个采样和第 i 帧隐写音频的第 j 个采样。

2. 以图像载体为例

对于图像来说,透明性可用隐写对象的视觉质量来刻画。视觉质量可采用主观评价法和客观评价法。主观方法就是以人的主观意念来评价图像质量的方法,具有代表性的主观评价方法是类似于表 2.1 的主观质量评分法,如表 2.2 所示。它有两类度量尺度:绝对性尺度和比较性尺度。评分的一般步骤如下:① 用某些原始标准图像建立质量等级标准;② 由观察者观看被评价的图像,并与图像质量等级标准作比较,得出被评价图像的等级;③ 对观察者的打分进行归一化平均。主观评价方法往往在很大程度上受观察者本身的知识背景、情绪、动机以及疲劳程度等因素的影响,评价结果的一致性较差,且从工程角度来看,该方法过于费时费力。因此,在图像处理中,传统主观评价方法都不能满足图像质量评价的要求。

表 2.2 图像质量的绝对性和比较性尺度

度量 \ 分值	5	4	3	2	1
绝对性评分	优	好	中	差	劣
比较性评分	一群中最好的	好于平均水平	该群平均水平	差于平均水平	一群中最差的

在客观评价方法中,最常用的评价指标是尖峰信噪比 (Peak Signal to Noise Ratio, PSNR)。在计算 PSNR 时,必须先计算出隐写图像与载体图像间的均方误差 (Mean Square Error, MSE),其公式为

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (c_i - s_i)^2 \quad (2.2)$$

其中, c_i 和 s_i 分别表示载体图像的第 i 个像素和隐写图像的第 i 个像素, N 为总像素个数。接着再利用 MSE 求出隐写图像的 PSNR 值如下

$$\text{PSNR} = 10 \lg \left(\frac{255^2}{\text{MSE}} \right) \quad (2.3)$$

单位为分贝 (dB)。需要说明的是,式 (2.3) 中的分子反映了尖峰 (Peak) 的概念,它指的是像素最大值,如果以 8 位二进制数表示一个像素值,即像素最大值为 255,最小值为 0,则两个像素间可能的最大差值为 255。例如将一个像素由白色变成黑色,则像素

值由 255 变为 0，这两个像素的差异值的平方即为 $255^2=(255-0)^2$ 。一般而言，PSNR 值越大表示隐写图像与载体图像越相像。经验上来说，PSNR 值高于 28dB 以上，隐写图像质量都是可接受的。

2.3.2 秘密信息的正确恢复率（鲁棒性）

第 1 章已经提到，信息隐藏系统的另一个重要评价指标是鲁棒性。信息隐藏算法的鲁棒性是指即使伪装对象 s 受到一定的扰动，仍然能从中提取秘密信息 m 的能力。用于版权保护的数字水印算法需要具有很强的鲁棒性，相比之下隐写术中对鲁棒性的要求不高。但是，无论如何，由于隐写术通常借助多媒体作为载体对象，而多媒体数据往往要作有损压缩处理以节省存储空间和传输时间，信息在传输过程中也会受到一定的噪声干扰，因此一定的鲁棒性要求还是必须的。

为了弱化隐写术对鲁棒性的要求，本书将针对隐写术的鲁棒性指标的名称改为秘密信息的正确恢复率。若秘密信息 m 是 L 比特的二进制串 $\{m_1, m_2, \dots, m_L\}$ ，而提取出来的秘密信息为 $m'=\{m'_1, m'_2, \dots, m'_L\}$ ，则可借用误码率（Bit Error Rate, BER）概念来描述如下

$$\text{BER} = \frac{1}{L} \sum_{i=1}^L e_i, \quad e_i = \begin{cases} 1 & m_i \neq m'_i \\ 0 & m_i = m'_i \end{cases} \quad (2.4)$$

若秘密信息 m 是一幅 $A \times B$ 的二值图像水印 $w=\{w_{ij}\}_{A \times B}$ ，还常常可以采用归一化相关系数（Normalized Correlation Coefficients, NC）来对原始水印 w 与提取的水印 w' 的相似性进行客观评价，定义为：

$$\text{NC} = \frac{\sum_{i=1}^A \sum_{j=1}^B (w_{ij} \cdot w'_{ij})}{\sqrt{\sum_{i=1}^A \sum_{j=1}^B w_{ij}^2} \cdot \sqrt{\sum_{i=1}^A \sum_{j=1}^B w'^2_{ij}}} \quad (2.5)$$

2.3.3 隐写容量

一般而言，在保证不可感知的条件下，隐藏的信息越多，鲁棒性就越差，因此每一个具体的隐藏方案都需要在不可感知性、鲁棒性和隐藏信息量之间进行折衷考虑。**隐写容量**（Stego Capacity）是指在保证不可感知条件下在给定的载体对象中能够隐藏秘密信息的最大信息量，以二进制数的位数为单位。在隐写系统中，如果将载体对象视为秘密信息的隐写信道，则隐写容量就相当于该隐写信道的信道容量。有鉴于此，多数关于隐写容量方面的研究采用信息论方法来求在一定约束条件下的平均互信息量最大值。

在目前的隐写容量分析中，最具代表性的是 **Moulin 模型**^[12]。Moulin 模型认为：隐写者的目的是找到一种合适的隐写方案使隐写系统能够安全、可靠传输的信息量最大，而攻击者的目的则是试图找到一种使隐写系统传输尽量少信息量的攻击方案，然后实施攻击。因此，隐写容量问题本质上是隐写者和攻击者之间的对策问题。如果将隐写过程看成隐写信道，**主动攻击**（Active Attack）看成攻击信道，载体对象和隐写密钥看成信道的边信息，那么隐写通信系统可用一条带边信息的串联信道表示，如图 2.5 所示。因此，Moulin 模型是隐写者和攻击者之间的博弈模型，费用函数是攻击信道的输入与输出间的互信息，攻击者试图使费用最小而隐写者力求使费用最大，该互信息的上限即为隐写容量。这样，Moulin 定义的隐写容量 Ω 可以描述为：

$$\Omega = \max_Q \min_A [I(u; y | k) - I(u; c | k)] \quad (2.6)$$

其中, Q 、 A 代表所有可能的隐写信道的集合和攻击信道的集合, u 为辅助随机变量, y 为攻击信道的输出随机变量, c 为表示载体对象的随机变量。符号 $I(u; v)$ 表示随机变量 u 和 v 之间的互信息, 符号 $I(u; v|w)$ 表示在 w 条件下随机变量 u 和 v 之间的条件互信息。

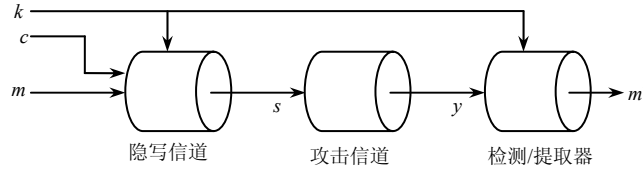


图 2.5 隐写通信系统的串联信道模型

2.3.4 安全性

隐写术的安全性就是隐写算法的抗攻击能力, 即隐写对象在承受一定程度的人为攻击的情况下而不致使秘密信息遭到破坏的能力。一个安全的隐写系统首先要求隐藏的秘密信息内容应是安全的, 应经过某种加密后再隐藏, 然后要求隐藏的具体位置也应是安全的, 至少不会因格式变换而遭到破坏。其次, 要求第三方在不知道隐藏算法和密钥时, 均不能获得秘密信息。与密码术一样, 隐写术也应把对秘密信息的保护转化为对密钥的保护, 因而密码术中对密钥的基本要求也适用于隐写术, 如必须要有足够大的密钥空间等。在设计一个隐写系统时, 密钥的产生、发放、管理等也都需综合考虑。

对于隐写的威胁来自**被动攻击** (Passive Attack) 和**主动攻击**两个方面。被动攻击是指攻击者对秘密信息的检测和提取, 主动攻击是指攻击者恶意篡改或伪装秘密信息, 破坏通信双方的通信。在被动攻击中, 提取技术基于检测, 而检测包括感知检测和统计检测, 攻击者通过量化秘密消息嵌入前后载体感知特征和统计特征的变化来验证秘密消息的存在性。隐写的本质就是保证秘密信息嵌入载体后不被隐写分析者成功检测, 因此, 隐写术的理论上安全性主要指对被动攻击者的检测的免疫性能。

在文献[13]中, Cachin 提出隐写系统安全性的信息论定义。在描述隐写的经典囚徒模型中, 如果只考虑被动攻击, 看守的任务是判别经过的消息是载体对象 c 还是隐写对象 s 。从信息论的观点来看, 隐写系统的安全性可通过载体对象 c 与隐写对象 s 之间的鉴别信息来度量。设载体对象 $c \in C$ 服从分布 P_C , 隐写对象 $s \in C$ 服从分布 P_S , 则隐写系统的安全性可用载体对象和隐写对象分布之间的鉴别信息 $d(P_C \| P_S)$ (相对熵) 来进行度量。如果 $d(P_C \| P_S) \leq \varepsilon$, 则称此隐写系统面临被动攻击时是 ε -安全的; 如果 $d(P_C \| P_S) = 0$, 则称此隐写系统是完全安全的。载体对象和隐写对象分布之间的鉴别信息 $d(P_C \| P_S)$ 定义如下

$$d(P_C \| P_S) = \sum_{c \in C} P_C(c) \log \frac{P_C(c)}{P_S(c)} \quad (2.7)$$

2.3.5 系统复杂度

系统复杂度反映的是秘密信息的嵌入和提取过程中的计算复杂程度。该指标对系统采用什么算法和处理器有很大影响, 并用于决定采用什么规格的通用 CPU 还是 DSP 芯片作为处理器。该指标的参数通常用单位时间内进行的乘、加运算次数和存储量等来描述。

2.4 基于文本载体的隐写术

2.4.1 引言

基于文本载体的隐写术，是通过改变文本模式或改变文本的某些基本特征来实现信息嵌入的方法，它使文本文档产生一定的变化，但是人的视觉对这种变化是不易察觉的。众所周知，信息隐藏通常是利用载体对象中的冗余信息的存在来工作的。由于人类视觉和听觉系统对某些信息不敏感，图像、声音、视频文件天然地包含噪声形式的冗余，所以针对这些载体的信息隐藏比较容易实现。但是，在文本里隐藏信息则比较困难，因为文本文件是直接对文字数据进行编码而成的，几乎不存在数据冗余，不可能通过修改原文件的有效数据来进行信息隐藏，而必须寻找那些不易引起视觉感知的方法。用于文本信息隐藏的文本载体类型^[14]可以分成纯文本文档、格式化文本文档、文本图像和纸质文档。

1. 纯文本文档

指 ACSII 码文本文档或计算机源代码文档，是所有文字文档中最为简单的文档。这种文档没有格式信息，编辑简单，使用方便，但很难嵌入秘密信息。

2. 格式化文档

一般指 Word、PDF、WPS、Postscript 等文档。这些文档中除了文本信息本身以外，还有很多用来标记文字格式和版面布局的冗余信息。对于这类文档，可以把隐藏信息嵌入到它们的文字的格式化编排中，例如行间距、字间距、字体、文字大小和颜色等不足以被人眼发现的微小变化中。

3. 文本图像

包括包含文本内容的灰度图像或二值图像，其中以二值文本图像为主。二值图像中的像素只有“0”或“1”两种取值。针对这类文档，现有的隐藏技术就是改动图像中的个别像素来嵌入秘密信息，可以把它归结为图像信息隐藏。

4. 纸质文本

这类载体若要实现隐藏信息的自动提取，需要先对其进行数字化、文字和排版识别等步骤。由于文本文档数据量小、编码简洁，使得它被人们广为利用，并成为传递信息最常用的形式之一，因而利用文本数字载体来隐藏、传递秘密信息是一种理想、实用而有效的方法。文本数字载体最显著特点就是数据与内容的高度一致性。它直接对文字数据进行编码，冗余非常少，因此在文本文档中嵌入信息的挑战性更大。

以文本为载体的隐写术研究相对较少，这主要是由文本的特殊性决定的，因为文本中的冗余信息非常少。文本是当前通信的主要形式之一，在军事、政务、商务、网络出版等方面发挥重要的作用，因此基于文本的信息隐藏技术具有很大的研究价值。目前，在文本中隐藏数据主要是将信息直接编码到文本内容中去（利用语言的自然冗余性），或者将信息直接编码到文本格式中（比如调整字间距或行间距），或者利用人们通常不易察觉的标点和字体的改变等方法。本节对当前现有的文本隐写术进行总结并分类加以介绍。

2.4.2 基于文档格式微调的隐写术

基于文档格式微调的隐写术是通过文本文档的空间域变换来嵌入秘密信息，通过将秘密信息藏入版面布局信息或格式化编排中达到嵌入的目的。文档的空间域不仅包括文本的

字符、行、段落的结构布局，也包括了字符的形状和颜色。由人类视觉系统的特点可知，文本中文字的某些微小的布局、结构变化不足引起人眼的注意。基于此，许多学者提出了改变行间距和字间距的隐写算法。有的学者提出在 PostScript 文档中通过微调字符位置、形状来插入隐藏信息的方法；也有的学者根据人眼视觉不易察觉字符颜色 RGB 值的微小变化的特点，提出将信息嵌入到字符的颜色中。基于文档格式微调的隐写术的安全性主要靠空间的格式隐蔽来保证，而且只能应用于格式化文档中，如 Word、PDF、Postscript、WPS 等文档。这类算法的隐蔽性较好，但是在使用字处理软件时，很容易有意无意地破坏原始文档的格式信息，经过简单复制操作，重新录入等都很容易破坏或去除隐藏信息，因此，该算法的抗攻击性不强，鲁棒性较差。下面简要介绍三种典型文档格式微调隐写方法。

1. 行移

行移和字移方法是文档格式微调隐写方法中最典型的两种方法，它们可以归为移位编码方法，是利用相邻文本（行或者词等）的相对位置关系来隐藏信息的方法。

行移就是在文本的每一页中，每间隔一行轮流地嵌入秘密信息，但嵌入信息的行的相邻上下两行位置不动，作为参考，需嵌入信息的行根据秘密信息的比特流进行轻微的上移和下移。在移动过的一行中编码一个信息比特，如果这一行上移，则表示嵌入“1”，如果这一行下移，则嵌入为“0”。该编码技术具有较强的鲁棒性，即使经过多次复制，或对页面按某个伸缩因子进行多次缩放，嵌入的秘密信息也可以检测出来。

2. 字移

字移就是通过将文本某一行中的一个单词进行水平移位来嵌入秘密信息。通常是在嵌入过程中，将某一个单词左移或右移，而与其相邻的单词并不移动，这些不动的单词作为解码过程的参考位置。此种方法与行移隐写术的原理大致相同，都是通过移动来实现的，一个简单的例子如图 2.6 所示。相对而言，字移隐写术能够隐藏更多的比特，但抗攻击能力较行移方法要弱。

3. 修改字符颜色

基于人眼视觉对字符颜色 RGB 值的微小变化不易察觉的特点，通过改变文本颜色也可以嵌入秘密信息。例如，根据人眼对蓝色最不敏感的特性，有的学者提出通过修改文本中字符的蓝色成分来嵌入秘密信息；有的学者提出通过置换文本中字符 RGB 颜色值中 R、G、B 的低 4 位值来嵌入秘密信息。

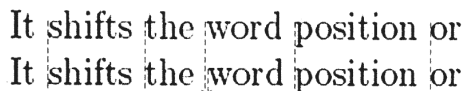


图 2.6 一个简单的基于字移的信息隐藏方法（通过对单词之间的距离微调来嵌入 1 比特信息）

2.4.3 基于空格和标点符号的隐写术

1. 基于添加空格的隐写术

基于空格的隐写术是利用人类视觉系统对标点符号和上下文之间的间隔、单词与单词之间的间隔不敏感的特点，通过在格式化文本中添加不足以引起人们注意的空格来嵌入隐藏信息的。为了使添加的空格不影响文本本身的含义，选择添加空格的位置是该类算法的核心问题。基于空格的隐写术在视觉上具有较好的隐蔽性，在实现上比较简单。但这种算法仅适用于英文文本，同时，这种编码方法有它自身的缺陷：一些文字处理软

件会自动插入或删除空格，从而破坏了嵌入的隐藏信息。

2. 基于标点符号的隐写术

在很多情况下，文本中的某些标点在使用上具有一定的任意性，对它们的误用不会严重影响文本的理解。例如短语“apple, banana, and orange”和“apple, banana and orange”都可以认为是正确的。另一方面，中文的逗号与英文的逗号，它们在文档中所占用的宽度是有一定差别的。在隐藏信息的时候，我们可以假设中文的逗号表示所要隐藏的信息位为 1，而英文的逗号表示隐藏的信息位为 0；或者相反。由于逗号在中文文档和英文文档中的使用比较频繁，所以利用逗号的变换可以隐藏更多的信息，但是这种方法存在的问题是标点符号不一致可能会被细心的读者发现，而且标点符号改变的随意性可能会影响文本内容的清晰甚至含义，所以这类方法应当谨慎使用。

3. 综合使用空格和标点符号的隐写术

实际上，人类的视觉系统（HVS）对标点符号和上下文之间的间隔、字与字之间的间隔、中文与英文标点所占用的字符宽度并不敏感，可以利用这个特点来进行信息隐藏算法的设计。根据标点符号两边的空格及标点类型，每个标点符号可隐藏 3 个比特的信息，隐藏方法为：用 b_0 , b_1 , b_2 分别代表标点左边的空格、符号类型、符号右边的空格，当需隐藏比特为 0 时，代表无空格或标点类型不改变；否则，代表有空格或标点类型改变。为了不使隐藏后的文本表现出更多的异常，只选取中、英文标点中差异不太大的，,;:?!等几个标点符号。具体隐藏算法描述如下^[15]。

步骤 1：将待隐藏的秘密信息转换为 0、1 比特流；

步骤 2：对载体文本进行格式化，主要是将载体文本中标点符号根据载体文本的类型（中文或英文）全部替换为与之相符的类型，并将符号两边的空格删除（对于英文类型的载体文本，保留标点右边一个空格）；

步骤 3：在载体文本中往后搜索下一个标点符号；

步骤 4：读取 3 比特秘密信息 b_0 、 b_1 、 b_2 ；

步骤 5：if $b_0 b_1 b_2 = 1 0 0$ then 在标点的左边插入一个空格；if $b_0 b_1 b_2 = 0 1 0$ then 替换标点类型；if $b_0 b_1 b_2 = 1 1 0$ then 在标点的左边插入一个空格并替换标点类型；if $b_0 b_1 b_2 = 0 0 1$ then 在标点的右边插入一个空格；if $b_0 b_1 b_2 = 1 0 1$ then 在标点的两边各插入一个空格；if $b_0 b_1 b_2 = 0 1 1$ then 在标点的右边插入一个空格并替换标点类型；if $b_0 b_1 b_2 = 1 1 1$ then 在标点的两边各插入一个空格并替换标点类型；

步骤 6：若还有需隐藏的秘密信息，转步骤 3，否则退出。

图 2.7 给出了上述算法的一个具体实例。通过对比不难看出，原始载体文本和隐写文本差别极小，所以很难引起检测者的怀疑。从算法描述中可以看出，这种算法非常简单，实现复杂度小，易于推广使用。同时，由于该方法的提取过程无需原始文档作参考，因此使用起来非常方便。但是，若重新修改标点符号及其左右空格，会对解码产生影响。

2.4.4 基于字符特征的隐写术

基于字符特征的隐写术通过改变文档中某个字符的某一特殊特征来嵌入秘密信息。特征可以是字体，也可以是字符（如 b、d、h、k 等）中的垂直线的长度，也可以是某些字符的高度。汉字是一种颇具特色的文字，它的结构独特、字体多样。和英文相比，汉字中可插入标记的可辨认空间较大，因此实施文本信息隐藏有着更大的潜力，略微改变汉字的笔画以及结构都可以嵌入秘密信息。下面介绍三种典型的基于字符特征的隐写术。

待隐藏的信息为：001 011 100。

载体文本：
(若为中文)信息隐藏是信息安全研究领域最新热点之一，隐写术及隐写分析是信息隐藏的重要研究内容。本文首先综述了信息隐藏的基本概念和原理，进而介绍了基于文本的几种典型的信息隐藏方法，并分析了相应的优缺点。
(若为英文)Information hiding has been one of the latest research hotspots in the area of information security, steganography and steganalysis are important research content of information hiding. In this paper, we summarize the general concept and the principle for information hiding. Also we introduce some typical realization methods of steganography based on text document, then analyze the merits and defects of these methods.

隐写文本：
(对应中文)信息隐藏是信息安全研究领域最新热点之一，隐写术及隐写分析是信息隐藏的重要研究内容。本文首先综述了信息隐藏的基本概念和原理，进而介绍了基于文本的几种典型的信息隐藏方法，并分析了相应的优缺点。
(对应英文)Information hiding has been one of the latest research hotspots in the area of information security, steganography and steganalysis are important research content of information hiding. In this paper, we summarize the general concept and the principle for information hiding. Also we introduce some typical realization methods of steganography based on text document, then analyze the merits and defects of these methods.

图 2.7 基于空格和标点符号的隐写术示例（隐藏在三个逗号中）

1. 基于字体替换的隐写术

在现有的文字处理中，大多都支持许多种字体，仔细观察发现不少字体比较相似，视觉上难以分辨，例如宋体和新宋体、仿宋和楷体、Times New Roman 和 Arial Narrow、PMingLiu 和 Gramaond 视觉上分别比较相似，例子如图 2.8 所示。基于字体替换的隐写术就是通过修改文本中一些文字的字体来隐藏秘密信息。例如，在进行隐藏的时候，如果原来的文字字体是宋体，我们记要隐藏的信息为“0”，如果将其改变为新宋体，则记要隐藏的信息为“1”。

宋体：隐藏
新宋体：隐藏
仿宋：隐藏
楷体：隐藏
Times New Roman：steganalysis
Arial Narrow：steganalysis
PMingLiu：steganalysis
Gramaond：steganalysis

在进行基于字体替换的隐写过程时，可以以单字、单图 2.8 几种相近的字体示例
词、句子等作为基本单位，对不同基本单位进行隐写所得到的效果也是不同的，这可以根据鲁棒性和需隐藏的信息量进行调整。图 2.9 分别举一些用中文和英文的例子来说明采用字体来进行信息隐藏的方法，其中，中文以单字、英文以单词为载体基本单位。待隐藏的信息统一为：001011100。

中文例 1：以宋体代表隐藏信息位 0，以新宋体表示隐藏信息位 1。载体文本为宋体。
载体文本：信息隐藏是信息安全研究领域最新热点之一。
隐写文本：信息隐藏是信息安全研究领域最新热点之一。
中文例 2：以仿宋代表隐藏信息位 0，以楷体表示隐藏信息位 1。载体文本为仿宋。
载体文本：信息隐藏是信息安全研究领域最新热点之一。
隐写文本：信息隐藏是信息安全研究领域最新热点之一。
英文例 1：以 Times New Roman 代表 0，以 Arial Narrow 表示 1。载体文本为 Times New Roman。
载体文本：Information hiding has been one of the latest research hotspots in the area of information security.
隐写文本：Information hiding has been one of the latest research hotspots in the area of information security.
英文例 2：以 PMingLiu 代表隐藏 0，以 Gramaond 表示隐藏 1。载体文本为 PMingLiu。
载体文本：Information hiding has been one of the latest research hotspots in the area of information security.
隐写文本：Information hiding has been one of the latest research hotspots in the area of information security.

图 2.9 基于字体替换的隐写术示例

通过上面例子，发现利用字体替换来进行隐写的方法是可行的，其隐藏效果主要与字体的选择有关，并且有较好的隐藏效果。这种信息隐藏方法的缺陷是必须在通信之前，双方预定义好通信协议，以保证信息的正确传输；另外，隐藏算法及提取算法比较复杂。

2. 基于汉字笔划的隐藏算法

汉字的基本笔画包括横、竖、撇、捺、折及标点符号等，这些是组成汉字的基本笔

画。在对汉字的修改中,选取黑色像素区域的 45° 或 135° 方向,对最普遍的笔画撇、捺、点等进行修改来嵌入隐藏信息,因为人眼对这些方向的视觉不太敏感。这类算法的鲁棒性要好于文档结构微调算法,但必须使用专门的字体库,通用性较差,只适合于 PDF 这类文档。

3. 基于汉字结构的隐藏算法

汉字具有很好的结构性,一般由若干个偏旁、部首组成,或由若干个汉字也可组成一个汉字。这些组合型的汉字按其位置分可以分为左右型、左中右型、上下型、交叉型等。例如:“镨”字,可用偏旁“钅”与汉字“容”合并为一个字来代替“镨”字。从表面上看,存储的是两个汉字,但是显示出来是一个汉字,不经过仔细辨认,很难辨别开。在不需要另外增加汉字库的情况下,利用标准汉字库中的汉字通过合并即可得到与标准汉字库有区别的汉字,从而达到嵌入隐藏信息的目的。但是,这种算法经过仔细的辨别(如放大等操作)还是很容易被攻击者发现的。

2.4.5 基于自然语言的隐写术

上面介绍的各种算法均有局限性,有的安全性较差,有的提取信息时需要原始载体文本作为参照,有的灵活性不好。这些算法都不能实现一个普遍适用、健壮文本隐写系统,而且容易被攻击。在嵌入秘密信息时,如何满足既不易被察觉、又能灵活适用隐写、水印和指纹等方面的要求,一个自然的想法就是利用自然语言处理技术,把秘密信息嵌入到文本的语义当中去。早期的基于自然语言的文本隐写术以绝对同义词替换、上下文无关文法等较简单的理论作为工具,通过修改载体文本或者根据秘密信息产生类似自然语言的载体文本。这类研究的理论工具简单,系统复杂性不高,易于实现,但是这些算法的不可见性普遍不好,同义词替换后的句子可能不太符合通常的表达。2000 年以后,学者们开始利用自然语言处理理论为工具,挖掘自然语言的语法、语义特征以寻求鲁棒性、隐蔽性更好的算法。这类研究主要以词法变换、句法变换和语义变换作为基本的操纵载体文本的手段。从目前发表的研究论文情况来看,这类研究还都处于理论探索阶段;从适用范围上考察,基于自然语言的隐写术应当是最有发展前途的技术。以下对现有的基于自然语言的文本信息隐藏技术分别加以简要介绍。

1. 基于同义词替换的隐写术

基于同义词替换的隐写方法是通过选择载体文本中在某一同义词库中出现的词,如“互联网”与“因特网”,“我们”与“咱们”,“动画”与“卡通”,“take”与“carry”等,并根据一定的编码方式对这些词进行同义词、同音词替换,以此来嵌入隐藏信息。例如,“实验”表示 1,“试验”表示 0;“互联网”表示 1,“因特网”表示 0;“图像”表示 1,“图象”表示 0 等。词是携带秘密信息比特的实体,每个词携带秘密信息的容量是其同义词个数的对数。同义词分为绝对同义词和相对同义词,绝对同义词指不存在任何语义的区别,按照相应的编码直接替换就可以了,但是这类同义词非常稀少。大部分同义词之间属于相对同义词,相对同义词之间存在一定程度的语义区别。如果对相对同义词直接进行替换,可能引起语义的混乱或产生歧义,这样很容易引起攻击者的怀疑,同时破坏了载体文本的可读性。因此在进行同义词替换时,必须考虑同义词的上下文语境,但是目前的自然语言处理技术要根据上下文环境判同义还比较困难。隐藏信息的容量与同义词库的大小有关,同义词库越大,文件的隐写容量通常也越大。这类隐写术的

优点是具有很好的透明性，并且算法简单，缺点在于需要通信双方事先约定好码本，且鲁棒性相对较弱，一旦遭到攻击（如部分删除或篡改），会影响秘密信息的恢复。

2. 基于句法变换的隐写术

该隐写术是利用改变措辞和句子结构而不显著改变句子意思和语气来嵌入隐藏信息。它充分利用句法分析器、句法树库等自然语言处理技术的研究成果，将隐藏信息定义为句子语法分析树的函数，句子是携带隐藏信息比特的实体，每个句子携带隐藏信息的容量和与其语义等价的句型数量呈正相关。句法变换方法主要有：主动被动语态变换、主题语前置、宾语前置、主语后置、移动附加语的位置、加入形式主语、插入“透明短语”等方法。

3. 基于语义变换的隐写术

基于语义变换的隐写术主要是在对句子进行深层理解的基础上，对句子语义结构进行转换来嵌入秘密信息。语义分析常常建立在某种语法或理论体系上，如语义语法、格语法、语义网络、蒙格塔语法、范畴语法、概念依存理论。美国普渡大学的 Victor Raskin 等人在**本体语义**（Ontological Semantics）的基础上提出采用 TMR（Text Meaning Representation）树的方式对文本中的句子进行表达，通过对 TMR 树的操作实现对文本中的句子语义保留的修改，从而达到嵌入秘密信息的目的。对 TMR 树进行操作主要有以下三种方式：嫁接（Grafting）、剪枝（Pruning）和等价信息替换（Substitution）。嫁接主要是根据上下文的有关信息来进行操作的；剪枝是针对上下文中一些重复的信息来对句子进行修改的，如果某个概念在文章中的其他地方出现了，则可以对其进行剪枝处理，这样的处理不会影响文章的意思；等价信息替换的方法不同于同义词替换，这种方法中的等价信息主要来源于**事实数据库**（Fact Database），该数据库是本体语义中的一个静态资源。基于语义变换的隐写术利用本体论知识分析文本语义，在不改变文本语义的条件下调整自然语言句子的内容，有效地实现了对短小文本的信息隐藏，但是这种算法使用了本体论的相关理论，需要本体库等大量基础数据的支持，而本体库的获得本身是一个非常难的问题，因此很难具体实现。

4. 利用自然语言生成方法产生隐写文本

利用自然语言生成方法产生隐写文本进行信息隐藏是文本隐写术的另一种研究思路，它不利用现成的载体，而是在自己控制下生成隐写文本，隐写文本单纯为了传递秘密信息而生。NiceText 系统是这方面的一个例子，它依赖于一个大的代码字典，这个字典由分类的单词组成（以单词、类型对的形式进行存储）。由文本风格源选择独立于密码的类型序列，NiceText 用和字典表中相应类型匹配的代码选择单词，将密码转换成句子。由于自然语言处理技术不够成熟，经过 NiceText 系统生成的句子的词汇间的相关性还比较弱，句子之间没有自然、连贯、完整的意思。

5. 基于机器翻译的文本信息隐藏技术

还有的学者提出在机器翻译过程中嵌入秘密信息的算法，该算法主要利用自然语言翻译过程中产生的噪声来嵌入秘密信息。因为自然语言在翻译过程中，存在大量的差异性。例如，对于同一个句子，同义词使它有了多种不同的翻译版本，同时在自动化文档翻译过程中，时常出现一些可容忍的误差或错误，这些都为信息隐藏提供了空间。基于机器翻译的文本隐写术的基本过程为：发送方首先获得原始文字，原始文字不需要隐藏，然后发送者通过编码器把原始文字翻译成目标语言，编码器对于每一个句子都有多

种翻译方式,选择其中一种嵌入秘密信息,然后将翻译后的内容发送给接收者,接收者用相同的编码工具翻译原始文字,通过对比翻译后的内容,接收者可以重构字节流从而得到秘密信息。这种信息隐藏方式具有很好的透明性,很难被攻击者检测出来,因为难以精确地判断错误是由于插入了秘密信息引起的还是由于翻译软件错误造成的,但是该算法有它自身的局限性,即需要在两种不同自然语言之间隐藏信息。

2.4.6 基于变换域的隐写术

基于变换域的隐写术是将文本先进行特定的变换后再隐藏信息的方法。文本信号与图像、声音信号显著的不同在于它本身不存在任何冗余信息。所有的文字都是用编码来代表的(如 0~127 是 ASCII,128 以上是汉字编码),这些编码数字是以整数形式存在的,它们是数据与内容高度一致的。数字发生任何变化,都将引起相应文字的错误。借用以图像、声音为载体进行隐写的思路,可将文本载体进行冗余化处理,如将信息流进行某种变换(小波变换、离散傅里叶变换、离散余弦变换、离散哈希变换等),利用变换后数据存在的冗余来隐藏信息。基于变换域的隐写术关键在于冗余化变换的选择,变换不同,冗余化效果也不同。通过大量的试验发现:对文本载体的比特流进行小波变换,其冗余化的效果最好。

2.4.7 对比和总结

上面对目前常见的文本隐写术分别作了介绍,表 2.3 对这些算法的优缺点进行了详细的分析和比较。

表 2.3 现有的文本隐写术比较

算法种类		算法性能			
		透明性	鲁棒性	容 量	其 他
文档格式微调法	行移	较好	较差	0.5bit / 行文本	适用格式化文本和二值文本图像
	字移			1 bit / 单词	
	颜色特征			大于 1 bit / 字符	适用格式化文本
基于空格和标点符号	空格	较好	较差	取决于采用的编码方法	适用格式化文本
	标点	较差			
基于汉字特点		较好	可抵抗局部篡改和噪声干扰	比较大	只适用于二值文本图像
基于自然语言	同义词替换	较好	具有抗格式修改及格式转换的能力,可抵抗局部内容篡改	没有确定值	高度依赖自然语言处理,适合于英文
	句法变换				
	语义变换				
	直接产生隐写文本			比较大	
	机器翻译			比较大	不适用于文本水印

2.5 基于图像载体的隐写术

2.5.1 引言

1. 人类视觉特性

图像作为一种传输视觉信息的媒介,是通过人眼接收信息的。各种图像的变换、压

缩、噪声影响等,要衡量它们对图像的影响有多大,必须与人的视觉联系起来加以研究。因此,基于图像载体的信息隐藏就是基于人的视觉特性,利用人的视觉系统的一些视觉掩蔽效应来实现的。人的视觉特性受到外界条件的影响,在不同照度和背景下,人眼对相同图像数据有不同的视觉感受。一个图像数据完全相同的色块在不同的环境、不同的陪衬下会有不同的视觉效果,因此人眼对图像的视觉不是对图像每个像素逐一产生响应。每个像素的视觉感受还受到周围像素的影响,每个局部的视觉响应不但取决于这个局部的图像数据,还与其周围局部的图像数据,乃至整幅图像的数据有关,人眼视觉是对一幅图像产生的总体感受。研究表明,人的视觉系统是一个非线性系统,其视觉灵敏度是有限的,它对微小的变化反应不敏感;在图像平滑区、边缘区和纹理区对噪声敏感程度是各不相同的。因此,基于图像载体的信息隐藏通常利用视觉掩蔽效应将图像划分成不同的块,在不同的噪声敏感块中分别嵌入不同的信息量,从而既提高了秘密信息的隐蔽性,又保证了嵌入信息量的最大化。下面归纳了以下一些与隐写术相关的人类视觉特性。

(1) Weber 定律

设背景照度为 I ,在均匀背景下,人眼刚可识别的物体照度为: $I+\Delta I$,其中 $\Delta I \approx I \times 0.02$ 。

(2) 频率敏感性和纹理复杂性

人类心理视觉研究表明,人的视觉系统对平滑区的变化很敏感,视觉阈值较低,只能嵌入少量的信息。随着频率的增高,人眼的分辨率会迅速降低。因此,人眼对属于高频部分的图像边缘的亮度误差并不敏感,可以嵌入适当强度的秘密信息;而人眼对纹理区的灰度变化很不敏感,视觉阈值较高,可嵌入较多的信息。

(3) 亮度敏感性

亮度敏感性描述了在固定亮度背景下,人眼对信号的视觉感知效果,这取决于背景平均亮度和目标信号的亮度水平。由于人眼对亮度具有非线性特性,在背景亮的区域,人眼对灰度误差不敏感;而对于背景暗的区域的灰度误差较敏感,视觉阈值偏低。因此我们可以充分利用图像的局部特征,对秘密信息的嵌入强度进行调整。

(4) 对比度特性

对比度特性描述了在给定的亮度背景下,人眼对目标信号的视觉感知掩蔽特性,也就是说一个信号在另一信号存在的情况下的可觉察性,尤其是这两个信号具有相同的空间频率、取向和位置时,掩蔽特性最强。一般该特性可以分为自对比度掩蔽和邻域掩蔽两类,前者是指具有相同空间频率、取向和位置的信号的掩蔽效应,后者则是指由空间相邻像素导致的掩盖效应,其反映了人眼视觉对平滑区域失真敏感性强于复杂纹理区域的特性。

(5) 方向敏感性

人眼对不同角度的空间频率视觉信号的响应也不同,对在垂直和水平方向的频率具有较强的视觉响应,而在对角线方向的频率响应显著下降。

(6) 灰度敏感性

人眼对图像像素本身的不同灰度具有不同的敏感性,其中对中等灰度区最为敏感,而对高灰度区、低灰度区敏感度降低。

(7) 边缘敏感性

图像的边缘信息对视觉很重要,特别是边缘的位置信息。人眼容易感觉边缘位置的变化,而对于边缘附近的灰度误差,人眼却比较不敏感。

人眼的这些视觉掩蔽特性是一种局部效应,受诸多因素影响。具有不同局部特性的区域,在其不被人眼觉察的前提下,允许改变的信号强度不同,这就是信息隐藏的关切点。

2. 隐写方式

近年来,人们提出了许多不同种类的隐写方式,其中大部分都可看作替代系统。这类系统试图用秘密信息来替代载体对象中的冗余部分,其主要缺点是鲁棒性较弱。根据嵌入过程中载体对象的变化可将隐写方式分为六种类型。① 替代系统:用秘密信息来替代载体对象中的冗余部分;② 变换域技术:在信号的变换域中嵌入秘密信息;③ 扩频技术:从扩频通信中得到的想法;④ 统计方式:通过改变一个载体对象的一些统计特性将秘密信息嵌入;⑤ 转换技术:通过信号转换来存储秘密信息并从载体对象中测出偏差来进行解密;⑥ 载体对象产生法:通过产生一个安全通信的载体对象来将信息嵌入。

3. 分类

另一方面,图像隐写术根据嵌入域可以分为空域隐写术、变换域隐写术和压缩域隐写术三类。实际可采用的隐写技术有很多,例如:最不重要位(Least Significant Bits, LSB)替代技术、图像降质隐写术、基于调色板的图像隐写术、量化抖动隐写术、二值图像隐写术、利用未用的和保留的计算机系统空间的隐写术等。限于篇幅,这里只介绍常见的空域隐写术,基于离散余弦变换(DCT)、离散小波变换(DWT)和离散傅里叶变换(Discrete Fourier Transform, DFT)的变换域隐写术,以及基于 JPEG 图像的压缩域隐写术。不论是空域还是变换域,最典型的图像隐写术是基于 LSB 的隐写术,因为该类方法只改变图像的一部分空域像素或变换域系数的最不重要位,比较容易实现,并且嵌入信息后图像质量的变化从视觉角度看是无法察觉的,所以应用比较广泛。根据载体图像类型的不同,可以把该类方法具体分为三个子类。

(1) 载体图像是没有经过压缩处理的原始图像

针对没有经过压缩处理的原始图像,如 BMP(Bitmap)图像,其基本隐写思路是:直接对图像空域中的一部分像素值的 LSB 位进行改变。这部分像素可以顺序选取、随机选取或者按照一定的步长选取。用该类图像作为载体实现 LSB 嵌入,可以提供比较大的嵌入空间。嵌入秘密信息后的图像视觉质量也不会有明显的改变,但是它会改变图像的一些统计特性,例如直方图,所以不能抵御统计袭击。

(2) 载体图像是调色板图像

调色板图像的像素值是指向调色板的索引,如 GIF(Graphics Interchange Format)图像。典型隐写思路是:先对调色板进行排序,然后改变一部分像素值的 LSB 位。由于该类算法在嵌入信息前先对调色板进行排序,故很容易引起怀疑,从而受到攻击。

(3) 载体图像是 JPEG 图像

JPEG 是目前最常用的图像存储格式,在 Internet 上大量存在,用其作为载体不易引起怀疑,故越来越多的隐写系统选择 JPEG 图像作为载体图像。尽管它提供的嵌入容量不如以上两种图像格式大,但是因为比较安全,使其成为最流行的隐写载体。对 JPEG 图像来讲,嵌入并不是在空域进行的,而是先将其转换到 DCT 域,通过改变 DCT 系数的 LSB 位来嵌入信息,如 JSteg、JPHide&Seek 以及 OutGuess 方法。

另外为了抵御统计袭击,现在的隐写算法在嵌入时都尽量保护图像的统计特性,例如下面要介绍的 Outguess 算法保留了大约一半的可用 DCT 系数,目的是用来纠正由于在另一半可用 DCT 系数的 LSB 位嵌入信息而引起的 DCT 系数直方图的变化。下面要介绍的 F5 算法是通过增加或减少 DCT 系数值来保持系数直方图看上去没有变化的。下面,我们分别介绍一些典型的空域隐写算法、变换域隐写算法和 JPEG 图像隐写算法。

2.5.2 空域隐写术

空域隐写算法是将秘密信息嵌入到 BMP、GIF 等图像像素或调色板数据的 LSB 上或是将秘密信息直接成块地嵌入在图像格式中。空域算法的优点是快捷、算法实现简单、隐写容量较大，但是由于大多数算法使用了图像最不重要的像素位，因此算法的鲁棒性较差，嵌入信息很容易受到滤波、图像量化、几何变形和加噪等操作的攻击。现在因特网上大多数隐写术软件都是基于空域隐写算法的，例如 **StegoDos**、**White Noise Storm** 和 **STools** 等。下面介绍几种典型的空域隐写算法。

1. LSB 替换隐写术

LSB 替换隐写术^[16]是最为简单的隐写术，它将秘密信息嵌入到载体 LSB 平面的某个子集中。通过顺序的或者随机的 LSB 替换，信息可嵌入到载体的 LSB 平面中。顺序 LSB 替换可以很方便地实现，但是有一个相对严重的安全问题，即在嵌入秘密信息后的载体中未修改的与已修改的部分有着明显的统计差异，而随机 LSB 替换则不同，信息能够随机地分散到载体中，大大提高了隐写的安全性。用加密过的秘密信息（比如秘密信息的伪随机排序）来替换这些 LSB 就更不会轻易被攻击者检测到载体中的秘密信息。

对于 256 灰度的图像载体来说，每个像素的灰度值 $p \in \{0, 1, \dots, 255\}$ 。每一个像素可以用 8 比特的二进制数来表示，从高位到低位排依次是 $b_7, b_6, \dots, b_1, b_0$ ，其中 b_0 就是最不重要位（LSB）。把所有像素的 $b_i, i \in \{0, 1, \dots, 7\}$ 位抽出来就构成了 b_i 位平面，如所有的 b_0 位构成了最低位平面，所有的 b_7 位构成了最高位平面。对于一幅自然图像，相邻像素灰度值的差别往往不大，即相邻灰度值的相关性比较强。而且位平面越高，对灰度值的贡献就越大，相邻比特的相关性也就越强，而最低位平面则类似于随机噪声。如图 2.10 所示，图 2.10 (a) 是 256×256 大小的灰度图像 Lena，图 2.10 (b) ~ (i) 则是 Lena 图像对应的 8 个位平面，图 2.10 (j) 是用一幅随机二值图像替换 LSB 位平面的结果，图 2.10 (k) 是用两幅随机二值图像替换最低两个 LSB 位平面的结果。由此可见，对于灰度图像，人眼不能分辨全部的 256 个灰度等级，LSB 替换法实际上就是利用人眼对图像的视觉冗余来替换最不重要位。显然，替换的位数越多，对载体图像的改变就越大，不可见性就越差。实验表明，替换最不重要位不会对图像视觉质量产生影响，甚至替换最不重要的几位（可达 4 位）都不会对图像视觉质量产生太大影响。

LSB 替换隐写术的一般嵌入过程可以描述为：记待嵌入的秘密消息比特序列为 $m = \{m_1, m_2, \dots, m_L\}$ ，其中 L 为比特序列长度；记载体图像像素集合 $c = \{c_1, c_2, \dots, c_N\}$ ，其中 N 为像素个数。从集合 c 中选择大小为 L 的一个子集 $c_s = \{c_{i_1}, c_{i_2}, \dots, c_{i_L}\}$ ($L \leq N, 1 \leq i_l \leq N, 1 \leq l \leq L$)，并对所有的 $i_l, 1 \leq l \leq L$ 作替换运算 $\text{LSB}(c_{i_l}) = m_l$ ，其中， $\text{LSB}(x)$ 表示 x 的最低有效位，从而把部分载体像素的 LSB 用秘密信息位代替。选择子集的方法可以是顺序法和随机法。进一步来看，假设原始像素值为 $p \in \{0, 255\}$ ，待嵌秘密信息位为 $b \in \{0, 1\}$ ，嵌入信息后的像素值为 p' ，那么 LSB 替换操作可以表达为

$$p' = \begin{cases} p + b & p \text{ 为偶数} \\ p - 1 + b & p \text{ 为奇数} \end{cases} \quad (2.8)$$

也就是说，如果 p 是偶数像素且负载的秘密信息位为 1 则进行加 1 操作，否则不变。反之，如果像素值 p 为奇数且负载的秘密信息位为 0，则进行减 1 操作，否则不变。LSB 隐写过程实质就是根据待嵌秘密信息位在像素值 $2i$ 和 $2i+1$ 中二选一的过程。

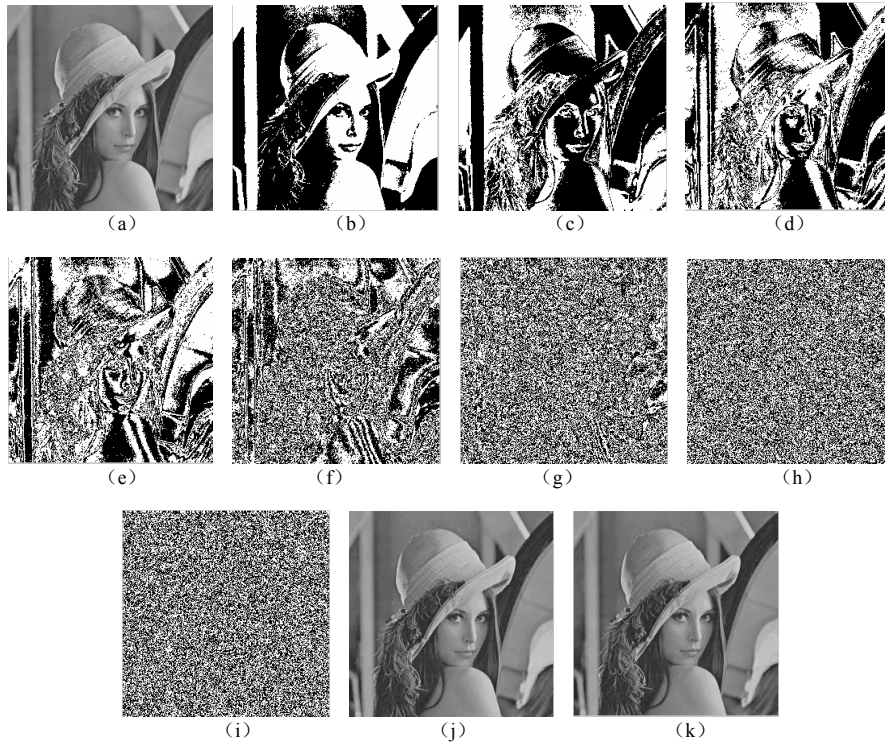


图 2.10 LSB 替换隐写术示例

下面举个简单的例子来说明顺序 LSB 替换隐写术和随机 LSB 替换隐写术的区别。假设要隐藏秘密信息序列“0110”到某图像的一组像素值中。原始像素值序列为：46 48 21 35 65 245 238 241 212 23。对应的 8 位二进制数序列为

```
00101110  00110000  00010101  00100011  01000001
11110101  11101110  11110001  11010100  00010111
```

若采用顺序嵌入方式，则选择的被嵌像素为（46 48 21 35），则嵌入秘密信息后对应二进制数序列就变成

```
00101110  00110001  00010101  00100010  01000001
11110101  11101110  11110001  11010100  00010111
```

即嵌入信息后的像素值序列就变成了：46 49 21 34 65 245 238 241 212 23。若采用随机嵌入方式，假如被选择的像素为（21 65 238 241），则嵌入秘密信息后像素值对应的二进制序列变成

```
00101110  00110000  00010100  00100011  01000001
11110101  11101111  11110000  11010100  00010111
```

则嵌入秘密信息后的像素值就变成了：46 49 20 34 65 245 239 240 212 23。

2. LSB 匹配隐写术

上述传统 LSB 替换隐写术很容易产生灰度直方图“值对”（Pair of Values）现象。这样一来，攻击者能够使用统计分析的方法成功地判断图像中是否含有秘密信息。为了应对隐写分析的攻击，Sharp 又提出了 LSB 匹配（LSB Matching）隐写术^[17]来弥补替换算法的不足。LSB 匹配隐写又常被称为±1 隐写，其嵌入秘密信息的原理是：当嵌入的秘密信息比特值与像素值最低位相同时，像素值不变；不相同，随机选择加 1 或减 1。

于是加 1 和减 1 两种转换以相等的概率出现, 减少灰度“值对”的出现, 而引起的图像失真却不变。该方法中关键点是随机选择加 1 或减 1 所采用的方法。通常, 可以采用一个随机数来控制。设 r 是一个 $[-1,1]$ 区间均匀分布的独立同分布随机变量, 假设原始像素值为 $p \in \{0,255\}$, 待嵌秘密信息位为 $b \in \{0,1\}$, 嵌入信息后的像素值为 p' , 那么 LSB 匹配隐写操作可以表达为

$$p' = \begin{cases} p & p \text{ 与 } b \text{ 同为偶数或同为奇数} \\ p+1 & p \text{ 和 } b \text{ 中一个为奇一个为偶且 } r > 0 \\ p-1 & p \text{ 和 } b \text{ 中一个为奇一个为偶且 } r < 0 \end{cases} \quad (2.9)$$

需要注意的是, 对于 $p=0$, 如果式 (2.9) 要求减 1, 则需要改为加 1; 而对于 $p=255$ 的情况, 如果式 (2.9) 要求加 1, 则需要改为减 1。

需要注意, 上述针对单个位平面的 LSB 替换隐写算法和匹配隐写算法都可以推广到使用最低第二、第三的位平面进行数据嵌入, 如使用两个最低位平面, 通常被称为 LTSB 或 L2SB (Least Two Significant Bits) 隐写算法, 但若对高位平面采用类似策略进行数据嵌入, 则会引起图像质量严重下降。

3. 位平面复杂度分割隐写术

Kawaguchi 和 Eason 在 1998 年提出了位平面复杂度分割隐写术 (Bit-Plane Complexity Segmentation Steganography, BPCS) [18], 其方法是首先将普通表示的位平面通过线性变换映射到特殊定义的位平面, 然后将各位平面划分成相同大小 (如 8×8) 的子块, 然后为每个子块计算复杂度, 将秘密信息隐藏在复杂度较高的位平面子块中。在不引起人眼可察觉的失真的前提下, BPCS 隐写术能将数据嵌入到较高的位平面, 使得最高数据嵌入容量增大, 嵌入率可以超过 1bpp (bits per pixel)。这种隐写术适用于位图、调色板图像和基于小波域表示的 JPEG2000 图像等。BPCS 隐写算法具有较好的隐蔽性和较大的嵌入容量, 其在空域和变换域中的嵌入原理和方法都相同。BPCS 应用于图像空域的具体实现方法如下所述。

(1) 对载体图像进行循环码编码 (格雷码), 编码后分成 8 个位平面, 将所有位平面分成相同大小的小块, 如 8×8 小块。

(2) 计算每个小块的复杂度, 其定义为所有相邻像素对中取值不等 (即一个为 0, 另一个为 1) 的像素对数目, 复杂度的最大值记为 F_{\max} 。对于 8×8 小块, 复杂度的取值范围为 0~112 的整数。

(3) 将复杂度大于 $\alpha \times F_{\max}$ 的载体位平面小块用于嵌入秘密信息, 这里 α 是系统参数, 其值要小于 0.5。 α 取得越小, 可嵌入的秘密信息量就越多, 实现时一般取 $\alpha=0.4$ 。

(4) 为了保证嵌入秘密信息小块后对载体位平面小块复杂度的改变不至于太大, 需要事先把秘密小块的复杂度都调到大于 $\alpha \times F_{\max}$ 。对于秘密信息组成的位平面小块, 如果其复杂度大于 $\alpha \times F_{\max}$, 直接替换载体位平面小块; 如果其复杂度小于或等于 $\alpha \times F_{\max}$, 则要对秘密小块作共轭处理 (Conjugation, 处理方法详见文献[18]), 用共轭处理后的新小块 (其复杂度肯定大于 $\alpha \times F_{\max}$) 替换载体位平面小块即可。显然, 需要一个共轭处理地图来记录受到共轭处理的秘密小块位置。

(5) 把所有图像块经过循环码译码回二进制形式并重组成隐写图像。

这里, 按格雷码形式划分位平面的原因是: 如果按二进制形式划分, 那么会有许多小块的复杂度大于 $0.5 \times F_{\max}$, 而 BPCS 隐写要求 $\alpha < 0.5$, 这样 BPCS 隐写将改变许多小块, 从而引起的失真较大, 而经过格雷码编码后, 许多位平面小块的复杂度将小于 $0.5 \times$

F_{\max} , 这样便可以通过调节 α 值来设置信息的不可感知性和嵌入容量。

4. 扩频图像隐写术

Marvel 等详细地阐述了基于扩频技术的隐写方法并提出了相关的检测技术^[19], 隐藏的消息能在没有原始图像参与但有嵌入密钥的条件下得到恢复提取。**扩频图像隐写术** (Spread Spectrum Image Steganography, SSIS) 的主要思想是通过在载体图像上叠加一个扩频调制的随机噪声来完成隐写过程。它增强了算法的鲁棒性, 但隐写容量却减小了。Marvel 等的 SSIS 理论清晰而明确, 严格地阐述了扩频技术的嵌入和提取的思路。SSIS 包含三大技术: 图像复原技术、错误控制编码技术和扩展频谱技术, 核心是扩展频谱技术。图像复原技术是从隐写图像中估计复原出原始载体图像。错误控制编码技术是为了解决图像复原误差。扩频隐写技术目的是为了使隐写引入的失真类似随机噪声, 主要方法是将二进制秘密信息与高斯白噪声序列进行调制扩频, 形成类似噪声的调制信号。

设载体图像为 c , 待嵌入的秘密消息序列为 $m=\{m_1, m_2, \dots\}$, $m_i \in \{0, 1\}$, 则 SSIS 的发送端或编码器的具体步骤如下: ① 基于密钥 k_1 对 m 进行加密处理并进行低比特率纠错编码 (目的是在检测端能尽量纠正传输过程可能发生的错误), 得到 $u=\{u_1, u_2, \dots\}$, $u_i \in \{0, 1\}$; ② 基于密钥 k_2 产生伪随机序列 $r=\{r_1, r_2, \dots\}$, $r_i \in N(0, \sigma^2)$; ③ 生成调制后的信号: $v=\{v_1, v_2, \dots\}$, $v_i = \text{Modulation}(r_i, u_i)$, 通常是对 u_i 进行 $\{0, 1\}$ 取值到 $\{-1, 1\}$ 取值的映射后与 r_i 直接相乘; ④ 基于密钥 k_3 对调制后的信号 v 进行数据组间交织操作 (这里数据分组是按照纠错编码来分的), 原因是纠错编码对每组有纠错能力上限, 而加入组间交织操作可以分散可能出现的错误, 从而得到 g ; ⑤ 得到伪装图像 $s=Q(c+g)$, 其中 Q 是量化操作, 目的是避免 $c+g$ 超出图像像素的取值范围。

在接收端或解码端, 对于合法接收者而言密钥 k_1 、 k_2 和 k_3 是已知的, 设收到的可疑伪装图像 s' , 则相应的秘密信息提取过程如下: ① 对 s' 进行图像复原滤波生成一个载体图像估计版本 c' , 然后计算 $g'=s'-c'$; ② 基于密钥 k_3 , 对 g' 进行解交织操作, 得到 v' ; ③ 基于密钥 k_2 生成与嵌入过程一样的伪随机序列 r ; ④ 执行解调操作 $u'_i = \text{Demodulation}(v'_i, r_i)$, 通常是 v'_i 除以 r_i 后取符号, 若符号为正则 $u'_i=1$ 否则 $u'_i=0$, 由此得到 u' ; ⑤ 对 u' 进行低比特率纠错编码相应的解码操作和基于密钥 k_1 的解密操作即可得到最终的秘密消息 m' 。有关扩频嵌入思想在后面 2.6.2 节中还将具体介绍。

5. 像素差隐写术

Wu 和 Tsai 于 2003 年提出了**像素差** (Pixel Value Differencing, PVD) 隐写算法^[20]。在该算法中, 载体图像 $c=\{c_1, c_2, \dots, c_N\}$ 被分割成互不交叠的小块, 每个小块由两个相邻像素组成。设第 i 个小块的两个相邻像素为 c_{2i-1} 和 c_{2i} , $i=1, 2, \dots, N/2$, 则秘密信息将隐藏在差分值 $d_i=c_{2i}-c_{2i-1}$ 中。如果差分值 d_i 较大, 则可以嵌入较多的秘密信息位。显然, 对于 256 灰度图像来说, 差分值的绝对值 $|d_i|$ 的范围是 $[0, 255]$ 。将这个范围划分成 J 个区间, 每个区间的长度都等于 2 的幂, 第 j 个区间 ($j=1, 2, \dots, J$) 可以表示为 $[L_j, L_j+W_j-1]$, 其中 W_j 为第 j 个区间长度。这样一来, 若差值 d_i 的绝对值属于第 j 个区间, 则可以嵌入 $\log_2 W_j$ 比特的秘密信息。因此, 对于第 i 个差值 d_i , 若它的绝对值落在第 j 个区间, 则从秘密信息比特流 m 中截取 $\log_2 W_j$ 比特并换算成十进制数 x_i , 设嵌入信息后的差值变为 d'_i , 那么 PVD 隐写操作可描述为

$$d'_i = \begin{cases} L_j + x_i & d_i \geq 0 \\ -(L_j + x_i) & d_i < 0 \end{cases} \quad (2.10)$$

然后,按一定方式去修改像素 c_{2i} 和 c_{2i-1} ,使它们的差值等于 d_i 即可。这里需要注意,若没有办法使得修改后的像素 c_{2i} 和 c_{2i-1} 落在 $[0, 255]$ 范围内,则需要放弃该像素对的修改,而转向下一个像素对,直到所有秘密信息位嵌完为止。

针对 PVD 算法,使用差分直方图就可轻而易举地进行攻击。为此,Zhang 和 Wang 于 2004 年提出了**改进像素差**(Modified Pixel Value Differencing, MPVD)隐写算法^[21],采用动态划分区间代替固定划分区间,使得安全性得以改善。限于篇幅,在此不作讨论。

6. 预测编码隐写术

Yu 等人于 2005 年提出了**预测编码隐写术**(Predictive Coding Based Steganography, PCBS)^[22]。PCBS 与 PVD 相类似,它使用一个空域预测器对像素值进行预测,通过修改像素值,秘密信息顺序地嵌入到图像的各个预测误差中,然后对嵌入信息后的预测误差作熵编码,以同时达到图像压缩目的。如果某个预测误差较大,则该预测误差可以嵌入较多的秘密信息位。在 PCBS 算法中,使用的空域预测器是**改进型 MED 预测器**(MMED, Modified Median Edge Detector),它的预测公式为

$$\hat{p} = \begin{cases} \min\{a, b\} & c \geq \max\{a, b\} \\ \max\{a, b\} & c \leq \min\{a, b\} \\ 0.5(a + b) & otherwise \end{cases} \quad (2.11)$$

式中, p 为当前像素, \hat{p} 为 p 的预测值, a 表示 p 的左邻像素, b 表示 p 的上邻像素, c 为 p 的左上邻像素。像素的实际值 p 减去预测值 \hat{p} 就得到预测误差 $e = p - \hat{p}$,显然 $e \in [-255, 255]$ 。假设每个像素嵌入 h 位秘密信息,将这 h 位二进制数转换成十进制数 d ,那么嵌入 d 后的预测误差变为

$$e' = \begin{cases} e + d - e \bmod 2^h & e \geq 0 \\ e - d + |e| \bmod 2^h & e < 0 \end{cases} \quad (2.12)$$

最后,对 e' 进行哈夫曼编码,就达到了数据嵌入与图像压缩同时进行的目的。数据提取过程相对比较简单,首先进行哈夫曼解码得到 e' ,其次计算

$$d = |e'| \bmod 2^h \quad (2.13)$$

提取出嵌入的数据 d 后,再根据嵌入位数 h ,就恢复出原始秘密信息。最后,与嵌入过程一样,对图像进行相同的预测得到预测值 \hat{p} ,可计算得到解码图像的像素值如下

$$p' = \begin{cases} \hat{p} + e' & \hat{p} + e' \in [0, 255] \\ \hat{p} + e' - 2^h & \hat{p} + e' > 255 \\ \hat{p} + e' + 2^h & \hat{p} + e' < 0 \end{cases} \quad (2.14)$$

实际上,前面的 PVD 也可视为 PCBS 的一类,即利用两个相邻像素中的其中一个作为另一个像素的预测值。但不同的是,PVD 将改动值扩散在两个像素中,而 PCBS 将改动停留在当前像素内。若预测方式被攻击者知道,则 PCBS 也存在与 PVD 相同的缺陷。由于 PVD、MPVD 和 PCBS 隐写算法将信息隐藏在像素的差分或预测值中,可将它们归为一类,称为**改变差分或预测值型隐写术**。

7. 混合进制系统隐写术

Zhang 和 Wang 于 2005 年提出了**混合进制系统**(Multiple Base Notational System, MBNS)隐写术^[23]。MBNS 将二进制表示的秘密数据转换为不同进制基表示的符号,然后改变像素值,使其对进制基进行求模得出的余数与这些符号相同。在纹理复杂区域内

的像素对应较大的进制基，所以嵌入的符号所携带的信息量较大。这种方法所产生的图像具有较好的视觉性能，分析者也难从直方图或者差分直方图上察觉秘密信息的存在，其最高数据嵌入率可以超过 1bpp。设一幅载体图像的大小为 $A \times B$ ，记载体图像和隐写图像在位置 (i, j) 上的灰度值分别为 $p_c(i, j)$ 和 $p_s(i, j)$ ，其中 $i \in [1, A]$ ， $j \in [1, B]$ ，则 MBNS 隐写术的嵌入过程可以描述如下。

(1) 令 $p_s(i, 1) = p_c(i, 1)$ ， $i \in [1, A]$ ； $p_s(1, j) = p_c(1, j)$ ， $j \in [1, B]$ 。也就是说，嵌入前后保持图像的第一行和第一列像素不变。

(2) 利用嵌入密钥 k 产生一个从 $p_c(2, 2)$ 到 $p_c(A, B)$ 的嵌入路径，记为 H 。其中，嵌入路径可以由用户采用任意方法设定，但必须保证 $p_c(i, j)$ 在 H 中的位置在 $p_c(i-1, j)$ 、 $p_c(i, j-1)$ 和 $p_c(i-1, j-1)$ 之后。

(3) 将二进制的秘密信息分段，每段长为 l 比特（例如 $l=32$ ）。

(4) 按顺序取一段秘密信息，将其转化为十进制数 m 。

(5) 按 H 中指定的嵌入路径进行嵌入。对于当前待嵌入位置为 (i, j) 的像素，计算其对应的进制基

$$b(i, j) = \min \left\{ \left\lceil \frac{\sigma(i, j)}{\Delta} \right\rceil, 16 \right\} \quad (2.15)$$

其中， Δ 为常数（例如 $\Delta=0.75$ ）， $\lceil \bullet \rceil$ 为向上取整操作， $\sigma(i, j)$ 是由 $p_s(i-1, j)$ 、 $p_s(i, j-1)$ 和 $p_s(i-1, j-1)$ 所求得的标准差。

(6) 若 $b(i, j) \leq 1$ ，则令 $p_s(i, j) = p_c(i, j)$ ，然后回到步骤 5（即得到的进制基并不使用，该像素也不进行信息嵌入）。否则，计算对应需嵌入的符号

$$d(i, j) = m \bmod b(i, j) \quad (2.16)$$

且令隐写图像的像素值为

$$p_s(i, j) = \arg \min_{v \in [0, 255], v \bmod b(i, j) = d(i, j)} |v - p_c(i, j)| \quad (2.17)$$

(7) 若当前秘密信息段未被完全嵌入，则用 $\left\lfloor \frac{m}{b(i, j)} \right\rfloor$ 更新 m ，转步骤 5。否则，转到步骤 4，对下一段秘密信息进行嵌入，直至所有秘密信息段都被嵌入。

由上述步骤我们可以看出，利用式 (2.15) 可以求得需要的进制基，利用式 (2.16) 可以求得待嵌入的符号，利用式 (2.17) 可以使隐写图像的像素值对进制基所求的模值与待嵌入的符号相等且使隐写图像的失真最小。

在秘密信息接收方，根据收到的可疑隐写图像 $\{p_s'(i, j)\}$ ，MBNS 隐写术的秘密信息提取过程可以简单描述如下。

(1) 利用嵌入密钥 k 得到与发送方相同的数据嵌入路径 H 。

(2) 利用式 (2.15) 可以求得所有 $p_s'(i, j)$ 对应的进制基。

(3) 利用求模运算计算所有大于 1 的进制基所对应的符号（即像素值相对于进制基值所求得模值）

$$d(i, j) = p_s'(i, j) \bmod b(i, j) \quad (2.18)$$

(4) 利用大于 1 的进制基和对应的符号位及数据嵌入路径 H ，恢复秘密信息。

为了验证算法的效果，图 2.11 给出了 512×512 的 256 灰度“Man”图像的载体图像、隐写图像（ $\Delta=0.75$ ，PSNR=39.9dB，嵌入率 0.29，隐写容量 6.1×10^5 ）以及载体图像与隐写图像之间的差异图像（为方便显示，差异图像的灰度级增强了 20 倍）。可见，

隐写图像和载体图像难以用人眼区分。并且，在纹理复杂区域以及边缘区域（对应于差异图像的深色区域），MBNS 隐写算法嵌入了更多的信息。表 2.4 给出了不同的 Δ 条件下，算法性能的变化。可见，随着 Δ 的下降，数据嵌入率变大，但图像失真越大。

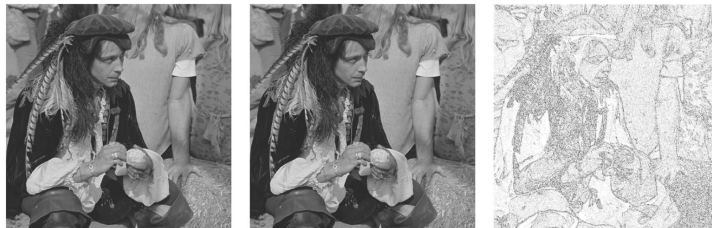


图 2.11 MBNS 隐写术作用下“Man”图像的载体图像、隐写图像和差异图像示例

表 2.4 MBNS 隐写术性能随着 Δ 的变化情况

Δ	1.50	1.25	1.00	0.75	0.50
隐写容量（比特）	4.2×10^5	4.6×10^5	5.2×10^5	6.1×10^5	7.4×10^5
嵌入率	0.20	0.22	0.25	0.29	0.35
PSNR (dB)	43.5	42.5	41.3	39.9	38.1

8. EzStego 调色板隐写术

在基于调色板的图像中，仅用特定色彩空间的一个颜色子集来对图像着色。每一个基于调色板的图像由两部分组成：一部分是调色板，定义了 L 种颜色 $\{v_0, v_1, \dots, v_{L-1}\}$ 的索引列表，为每一个颜色向量 v_i 分配一个索引 p_i ；另一部分是实际图像数据，它保存每一个像素的调色板索引，而不是保存实际的颜色值。如果整个图像仅使用一小部分颜色值，这种方法大大地减少了文件的尺寸。两种最流行的调色板图像格式是 GIF 和 BMP 格式。基于调色板的图像中有两种方法对秘密信息进行隐写：操作调色板或操作图像数据。如上面描述的替换方式一样，颜色向量的 LSB 也能用于秘密信息隐写。另外，因为调色板不需以任何方式排序，在以调色板保存颜色时，可选择不同排序方式来对传递秘密信息。根据排列组合，一共有 $L!$ 种不同方式对调色板进行排序，所以有足够的对一个秘密短信息进行编码。然而，所有使用调色板顺序保存信息的方法都不具有稳健性，任何攻击者都能简单地以不同方式排序调色板而破坏秘密信息（甚至从视觉上看不出图像的修改）。另外，还可以在图像数据中进行信息隐写，由于调色板上相邻的颜色值在感观并不接近，这样就不能简单地修改一些图像数据的 LSB。因此，为了使相邻颜色在感观上接近，在开始嵌入数据之前需要对调色板进行重新排序。

在所有调色板图像隐写算法中，由 Machado 提出来的 EzStego (<http://www.stego.com>) 是其中一种比较典型的方法，其思想是以秘密信息位替换载体图像像素在排序后的调色板对应索引的 LSB。EzStego 隐写术需要创建一个临时的重新排序后的调色板，它遵循最短路径原则。由于视觉对颜色辨别能力的限制，通常很难辨别排序后调色板上两个相邻颜色之间的差别。Ezstego 的嵌入是从图像的左上角到右下角连续逐行进行的，图 2.12 用简化的调色板阐述了 EzStego 的嵌入思想。对于原始载体图像中的第 i 行、第 j 列的像素 $p_c(i, j)$ （设索引值为 7），假设其在排序后的调色板中的索引为 4（二进制“100”），为了嵌入秘密消息位 1，则索引值将改为 5（二进制“101”），所得到的隐写像素 $p_s(i, j)$ 在原调色板中的索引值为 3，也就是原来位置上的像素 $p_c(i, j)=7$ 变为 $p_s(i, j)=3$ ；如果要嵌入的秘密消息位为 0，则什么也不变，也就是说原来的颜色不变， $p_s(i, j)=p_c(i, j)$ 。

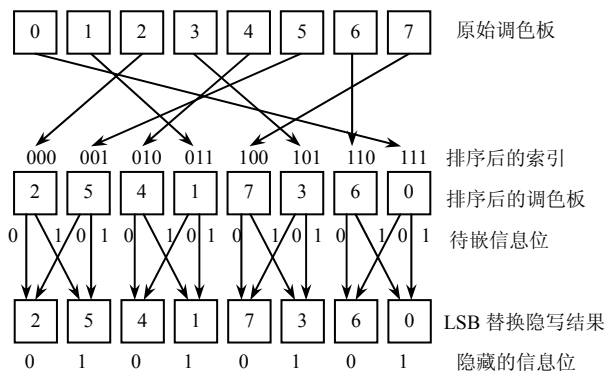


图 2.12 EzStego 隐写术嵌入原理示意图

2.5.3 变换域隐写术

变换域算法是在离散傅里叶变换 (DFT) 域、离散余弦变换 (DCT) 域、离散小波变换 (DWT) 域等上实现信息的 LSB 嵌入。算法主要是通过修改主信号某些指定的频域系数来嵌入数据。考虑到对低频区域系数的改动可能会影响到主信号的感知效果而高频系数又容易被破坏, 因此, 隐写术一般选取信号中频区域上的系数来嵌入秘密信息, 从而使之既满足不可感知性又满足对诸如失真压缩等操作的鲁棒性。下面分别介绍图像的三种正交变换和各自变换域下的典型隐写术。

1. 图像 DFT 域隐写术

离散傅里叶变换是线性系统分析的有力工具, 在数字信号处理技术中占有重要的地位。由于 DFT 是正交变换, 计算时可以采用快速算法。特别地, 信号的离散傅里叶变换系数有明确的物理含义, 因此它在通信、雷达、声纳、遥感、医学、图像处理、语音合成与分析等许多领域得到了广泛应用。DFT 是由法国科学家傅里叶 (公元 1768—1830) 所提出来的, 主要目的是利用 DFT 取得信号所对应的频谱, 再由频谱读取信号参数。其主要概念是将长度为 N 的离散时间信号表示成正弦波或者是余弦波的组合, 其定义如下

$$F(u) = \frac{1}{N} \sum_{x=0}^{N-1} f(x) e^{\left(\frac{-j2\pi ux}{N}\right)} \quad (2.19)$$

其中, x 表示采样位置, u 表示转换后的频率系数位置, $0 \leq u, x \leq N-1$, $f(x)$ 表示第 x 个采样值, $F(u)$ 表示第 u 个频率系数, 其中

$$e^{j\varphi} = \cos \varphi + j \sin \varphi \quad (2.20)$$

式 (2.19) 对应的离散傅里叶逆变换 (Inverse Discrete Fourier Transform, IDFT) 为

$$f(x) = F^{-1}(u) = \sum_{u=0}^{N-1} F(u) e^{\left(\frac{j2\pi ux}{N}\right)} \quad (2.21)$$

上面的方程式是针对一维数据的公式。对于大小为 $A \times B$ 的二维图像, 公式变为

$$F(u, v) = \frac{1}{A \times B} \sum_{x=0}^{A-1} \sum_{y=0}^{B-1} f(x, y) e^{-j2\pi \left(\frac{ux}{A} + \frac{vy}{B}\right)} \quad (2.22)$$

其中, (x, y) 表示二维像素的位置, (u, v) 表示变换后的频率系数位置, A 和 B 为图像的长度和宽度, $0 \leq u, x \leq A-1$, $0 \leq v, y \leq B-1$ 。其离散傅里叶逆变换为

$$f(x, y) = F^{-1}(u, v) = \frac{1}{A \times B} \sum_{u=0}^{A-1} \sum_{v=0}^{B-1} F(u, v) e^{j2\pi(\frac{ux}{A} + \frac{vy}{B})} \quad (2.23)$$

离散傅里叶变换在隐写术中也受到了高度重视。离散傅里叶变换是复数变换，在幅度和相位满足特定的条件下，秘密信息既可以嵌入到媒体信号的幅度上，也可以隐藏在它的相位中。但是，傅氏变换系数有其特殊点，使用时需要注意以下方面。首先，实信号的傅氏变换系数包含实部和虚部，因此隐藏信息时应考虑在实部、虚部、幅度还是相位中隐藏信息。另外，实信号的傅氏变换系数是对称的，因此，如果在傅氏变换的幅频特性中隐藏信息，应该保证不改变系数的对称性，这样才能保证傅氏逆变换后仍能得到实信号。

基于 DFT 变换的信息隐写的基本流程为：① 对于载体图像进行 DFT 变换；② 将变换后的幅值取整，并分解为一系列位平面表达；③ 将以二进制表示的秘密信息嵌入到幅值位平面中，一种较为简单的方式是使用秘密信息直接替换某个幅值位平面。由于 DFT 变换具有特殊的对称性，嵌入的秘密信息也应该满足相应的对称性，实际可以嵌入的信息量将少于幅值位平面所包含的元素个数。对于大小为 $N \times N$ （假定 N 为偶数）的载体图像矩阵 C ，通过 DFT 变换，得到频域矩阵 F ，并平移其频域原点至中心后，可将频域矩阵 F 用如下矩阵分块表示

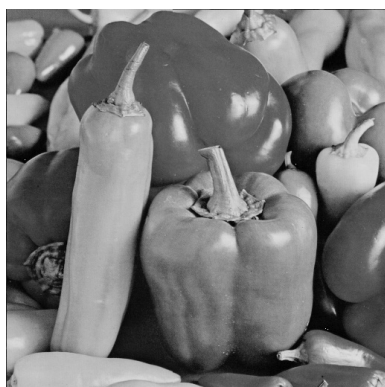
$$F_{N \times N} = \begin{bmatrix} U_{1 \times 1} & A_{1 \times (N-1)} \\ B_{(N-1) \times 1} & V_{(N-1) \times (N-1)} \end{bmatrix} \quad (2.24)$$

其中，频域原点位于 V 矩阵分块的中心，4 个矩阵分块均各自具有对称性。考虑其对称性，实际可嵌入的信息位数为 $1 + \{[(N-1)+1] / 2\} \times 2 + [(N-1) \times (N-1)+1] / 2 = N^2 / 2 + 2$ 。

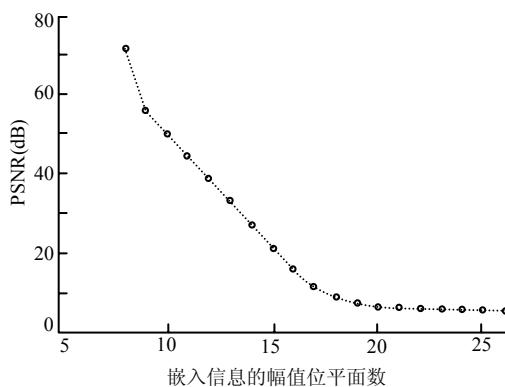
理论上，从隐藏有秘密信息的图像 DFT 变换的幅值位平面中可以恢复出嵌入的秘密信息。然而，当通过 DFT 逆变换得到伪装图像时，需要对 DFT 逆变换的实数值取整，并裁减至 0~255 范围，从而导致嵌入的信息和提取的信息可能存在微小差异，无法保证其完全相同。为了更好地恢复隐藏的秘密信息，可以对其进行纠错编码，并将纠错码嵌入到图像 DFT 变换的幅值位平面中。具体的纠错编码方法有很多，为简单起见，可采用 (7, 4) 汉明纠错码，它表示汉明纠错码的每一分组由 7 位组成，包括 4 位信息位和 3 位校验位，可保证任意 2 个编码的信息至少有 3 位不同，并可在仅有 1 位错误时予以更正。采用纠错编码嵌入秘密信息导致的不足减少了实际可隐藏的秘密信息量。

为了验证图像 DFT 域全频域隐写术的有效性，我们采用大小为 512×512 的 256 灰度 Peppers 图像为测试图像，如图 2.13 (a) 所示。分别在 DFT 变换全频域的不同幅值位平面嵌入相同的秘密信息，其最大幅值位平面个数为 26，考察嵌入信息的不可感知性及秘密信息的正确恢复率。其中嵌入信息的不可感知性使用嵌入秘密信息后图像相对于原始图像的峰值信噪比 PSNR 进行度量，秘密信息的正确恢复率使用最终恢复的秘密信息与原始嵌入的秘密信息之间的比率进行度量，实验结果如图 2.13 (b) 和图 2.13 (c) 所示。由图 2.13 (b) 可见，随着嵌入信息的幅值位平面数增多，其不可感知性逐步减小。为了保证嵌入信息的不可感知性，一般只将信息嵌入到中间及其以下的幅值位平面。对于第 1 至第 7 幅值位平面，嵌入信息后图像与原始图像完全相同，PSNR 的值为无穷大。图 2.13 (c) 分别列出了在第 8~18 幅值位平面不采用和采用纠错编码时秘密信息的正确恢复率。由图 2.13 (c) 可以得到如下结论，① 将秘密信息嵌入中间幅值位平面，可以在满足嵌入信息的不可感知性[图 2.13 (b)]的同时，达到较高的恢复率。② 在不采用纠错编码的情况下，选择合适的幅值位平面嵌入信息，也可达到较高的恢复率。例如：将未进行纠错编码的信息嵌入在第 13 幅值位平面，秘密信息的正确恢复率已达到 98.68%

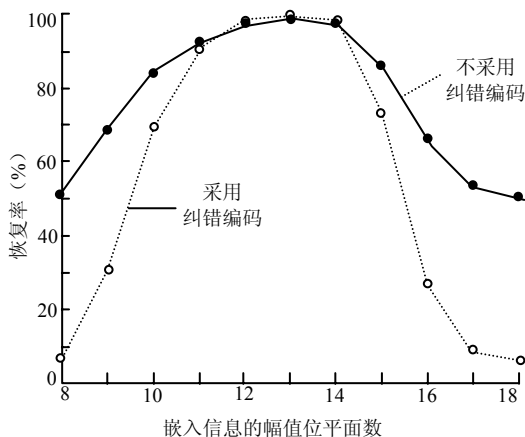
(使用不同的载体图像和嵌入信息时, 该结果稍有出入)。③ 在中间幅值位平面附近采用纠错编码嵌入信息, 优于不使用纠错码; 而在较低或较高的幅值位平面嵌入信息, 使用纠错码的恢复率反而较不使用纠错码差, 原因在于采用纠错码时, 对于一个 7 位分组而言, 当提取信息与嵌入信息相同或仅有 1 位不同时, 可通过纠错实现 100% 的恢复率; 若提取信息与嵌入信息有 2 位或 2 位以上不同, 则恢复率为 0。若不采用纠错码, 则不会出现这种对于一个 7 位序列要么完全恢复, 要么完全不恢复的情况, 而可能部分恢复。当嵌入信息在中间幅值位平面附近时, 提取信息与恢复信息大多相同, 在一个 7 位分组中出现 2 位及 2 位以上不同的情形很少, 因此采用纠错码恢复信息的效率高于不采用纠错码的。在较高或较低的幅值位平面嵌入信息, 由于提取信息和嵌入信息有更多差异, 使得更多的 7 位分组包含 2 位或 2 位以上不同, 导致使用纠错码的恢复性能反而下降。总体而言, 将经过纠错编码的秘密信息嵌入中间幅值位平面, 可实现不可感知性及恢复率的最佳组合。因此, 在使用 DFT 变换实现信息隐写时, 应将秘密信息通过纠错编码后嵌入到 DFT 变换的中间幅值位平面中。



(a) Peppers 载体图像



(b) 不可感知性



(c) 秘密信息的正确恢复率

图 2.13 图像 DFT 域全频域隐写术性能

上面考虑在图像 DFT 变换的全频域中隐藏信息, 没有考虑生成的隐写图像可能遭受图像压缩。由于 JPEG 压缩通常针对图像的高频部分, 故在全频域隐藏信息将严重影响 JPEG 压缩后隐藏信息的恢复率。为此, 可仅在 DFT 变换的低频域隐藏信息, 以实现一定程度上提高隐藏信息在 JPEG 压缩下的生存能力, 其代价是隐藏的信息量有所减少。

2. 图像 DCT 域隐写术

离散余弦变换 (DCT) 是 Ahmed 等人于 1974 年提出来的。因为其变换矩阵的基向量很近似于 Toeplitz 矩阵的特征向量, 而 Toeplitz 矩阵又体现了人类语言及图像信号的相关特性, 因此 DCT 常常被认为是语音和图像信号的准最佳变换, 其性能接近 KLT 变换。由于 VLSI 技术的发展, 结构有规律或易于并行的一些 DCT 算法已能用专用 IC 或微码实现, 这就更牢固地确立了 DCT 目前在图像编码中的重要地位。JPEG、H.26X、MPEG-1, MPEG-2 等图像压缩国际标准均是基于 DCT 的。对于大小为 $A \times B$ 的二维图像, DCT 是另一种将空间域像素转成频率域系数的技术, 其变换公式为

$$F(u, v) = \frac{2}{\sqrt{A \times B}} C(u) C(v) \sum_{x=0}^{A-1} \sum_{y=0}^{B-1} f(x, y) \cos \left[\frac{(2x+1)u\pi}{2A} \right] \cos \left[\frac{(2y+1)v\pi}{2B} \right] \quad (2.25)$$

其中, $C(u) = \begin{cases} \sqrt{\frac{1}{2}} & u=0 \\ 1 & \text{其他} \end{cases}$ 。与 DFT 一样, 离散余弦变换也有逆变换 (Inverse Discrete Cosine Transform, IDCT), 其公式为

$$f(x, y) = \frac{2}{\sqrt{A \times B}} \sum_{u=0}^{A-1} \sum_{v=0}^{B-1} C(u) C(v) F(u, v) \cos \left[\frac{(2x+1)u\pi}{2A} \right] \cos \left[\frac{(2y+1)v\pi}{2B} \right] \quad (2.26)$$

DCT 被使用在许多压缩技术上, 例如 JPEG 图像压缩、MPEG 视频压缩等。一般而言, 以 DCT 进行图像压缩时, 会先将图像切割成许多大小为 8×8 的区块, 再针对每个区块进行 DCT 变换。经过变换后, 系数值按之字形编排如图 2.14 (a) 所示, 其中 $F(0,0)$ 为最重要系数, 又称为 DC 值, 而其他系数为 AC 值。DCT 系数表现了图像的能量分布情况, 越靠近左上角的系数越重要, 我们称这些系数为低频系数。反之, 越靠近右下角的系数越不重要, 是为高频系数。而在中间的系数称为中频系数, 其分布图如图 2.14 (b) 所示。这里强调的“不重要”, 是因为高频系数做些微小的修改, 对整体图像质量的影响很小, 尤其对人类视觉系统而言, 不容易察觉出来。所以, 图像压缩技术多将高频系数量化, 减少所需记录的信息, 达成高压缩率的目的。

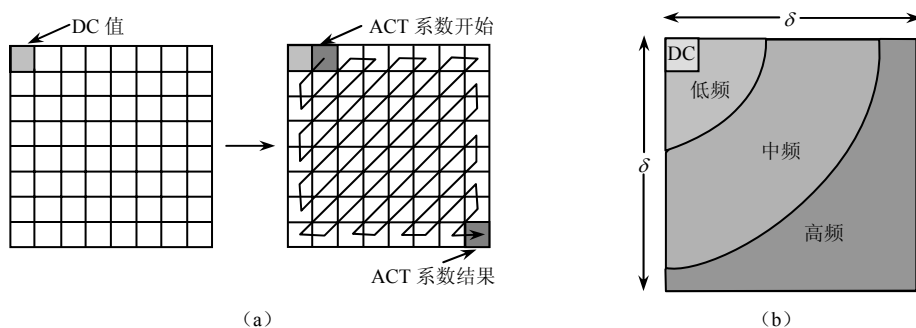


图 2.14 8×8 DCT 变换的频率域系数分布

图像 DCT 域的隐写术的基本思想与 DFT 域的类似。通常, 可以采用两种变换方式, 一种是针对整幅图像的 DCT 变换, 另一种采用与 JPEG 压缩标准兼容的 8×8 DCT 变换。目前大多数 DCT 域隐写算法都尽量与 JPEG 压缩兼容, 所以基本上采用 8×8 分块 DCT 变换。与 JPEG 压缩兼容的图像 DCT 域隐写术的嵌入和提取秘密信息有如下三种方案: ① 不进行 JPEG 压缩的 DCT 量化和反量化过程, 直接修改部分 DCT 系数, 隐

写图像还是非 JPEG 格式；② 进行 JPEG 压缩的 DCT 量化和反量化过程，直接修改量化后的部分 DCT 系数，隐写图像还是非 JPEG 格式；③ 输入和输出都是 JPEG 图像的隐写术，属于压缩域隐写术，这将在下一小节中介绍。本书所指的 DCT 域隐写术是指前两种方案或者针对整幅图像的 DCT 变换而言的隐写术。

在图像 DCT 域隐写术中，需要重点考虑的问题是 DCT 系数或 DCT 量化系数的选择问题。针对与 JPEG 压缩兼容的隐写方案，首先来看 **DC 分量**，即直流分量是否适合嵌入秘密信息。DC 分量是之字型排列的 0~63 个 DCT 系数中第 0 个 DCT 系数。DC 分量携带的感知能量最大，嵌入秘密信息后，隐写图像将产生强烈的块效应，不可感知性很差。然而秘密信息嵌入这个位置，在隐写图像受到有损压缩等攻击后，秘密信息受到的攻击很小，鲁棒性却较好，总之用 DCT 系数嵌入秘密信息时，不可感知性和鲁棒性是一对矛盾体，为了保证图像的质量，一般不用 DC 分量作为嵌入位置。再来看 **AC 分量**，即交流分量。AC 分量可分为低频带、中频带和高频带。大量事实证明，JPEG 压缩标准中的图像经过 DCT 变换后再经过量化后，之字形排列的 0~63 个 DCT 系数中，一般数值不为 0 的值多集中于左上角，而 DC 分量一般不为 0。低中频带中往往只有大约 5 个系数不为 0，且对图像能量的影响较小。中高频带大部分系数为 0。因此，兼顾嵌入秘密信息的不可感知性和鲁棒性，低中频带都可以考虑作为嵌入秘密信息的理想位置。中频带和高频带分量携带能量较少，从保证嵌入信息的不可感知性角度看是好选择，但在中高频带中嵌入秘密信息很容易被有损压缩等攻击去除，鲁棒性很差。若用于数字水印，则需要更加注重嵌入秘密信息的鲁棒性，通常可以利用低频带作为 DCT 域信息嵌入区域。若用于保密通信，则更注重隐藏信息的不可感知性以及隐藏信息容量，故可以利用低中频带作为 DCT 域信息嵌入区域，比如选择之字排列的 0~63 个 DCT 系数中序号为 3~12 的用来嵌入秘密信息。另外一种改进的方法可以考虑减小 DCT 变换后的量化步长，以增加 AC 分量中不为 0 的量化系数，用于嵌入秘密信息。

图像 DCT 域隐写术方法很多，目前主要分为三类。① 通过修改 DCT 系数的 LSB 来嵌入秘密信息；② 基于 DCT 中低频系数间相互关系来嵌入秘密信息；③ 借鉴小波域零树概念，基于 DCT 零树来嵌入秘密信息。限于篇幅，这里只介绍一种比较简单的由 Zho 和 Koch 两位学者于 1995 年时提出的基于系数关系的 DCT 域隐写方案^[24]：发送者将载体图像分成 8×8 的子块，在每一子块中调整两个 DCT 中频系数的相对大小来隐藏一个秘密信息位。在通信开始前，发送者和接收者必须约定使用的两个系数的位置，通常应该选用 DCT 中频系数以确保嵌入信息不容易因 JPEG 压缩而完全丢失。如果要隐藏信息的某子块的系数一大于系数二就代表嵌入信息位“1”，否则代表嵌入“0”。在嵌入阶段，如果这对系数的相对大小关系与要嵌入的比特不匹配，就相互交换这两个系数即可。最后，发送者执行逆 DCT 变换把频域系数变换回空间域，就得到隐写图像块，所有隐写图像块拼起来就是隐写图像。在提取信息时，对所有图像块进行 DCT 变换，通过比较每一块中的两个系数，就可以提取所隐藏的信息位。

3. 图像 DWT 域隐写术

离散小波变换（DWT）也是一项常用的变换技术，目前常被人使用的 JPEG2000 压缩标准就是使用 DWT 技术进行图像压缩的。小波变换是由法国数学家 Morlet 于 1980 年在分析地震资料时引入的。他与 Grossman 等首先提出了“小波”（Wavelet）概念，建立了完整的连续小波变换的几何体系，其基础是平移和伸缩变换下的不变性，这使得它能将一个信号分解成对空间和尺度（相当于频率）的独立贡献，同时保持原信号信息。因

此, 可以认为小波函数的伸缩平移系用于可测平方可积函数空间展开的概念是由他们首先提出的。小波级数理论依赖于小波基的发展, 1982 年 Stromberg 构造了第一个正交小波基。20 世纪 80 年代后期是小波发展的一个重要时期, 出现了大量的正交小波类。1988 年 Mallat 提出多分辨率分析概念, 为此前各种小波基的构造建立了统一的框架, 同时将离散小波变换与 Daubechies 紧支正交小波相结合提出了 Mallat 塔式分解算法, 为 DWT 建立了快速算法, 也促进了小波在信号处理中的应用。与 DCT 变换相比, DWT 变换弥补了 DCT 变换不适用于带宽较宽信号的不足。由于在 DWT 变换中, 图像是被分层而不是被分块进行处理的, 克服了 DCT 的块效应。另外从计算复杂性而言, DWT 比 DCT 低得多。限于篇幅, 有关小波变换的基础请读者自行去翻阅有关文献, 在此不做详细介绍。

为了便于大家理解, 这里介绍以 Haar 方法为基础的最简单和最常用的小波变换。在 Haar 方法中, 图像先经过水平处理, 如图 2.15 所示。图 2.15 (a) 中 A 、 B 、 C 、 D 分别表示四个不同的像素值, 以水平方向做处理, 分别计算 $A+B$ 、 $C+D$ 、 $A-B$ 及 $C-D$ 的结果, 如图 2.15 (b) 所示。接着将图 2.16 (a) 再进行垂直方向处理, 如图 2.16 (b) 所示, 分别计算 $W+X$ 、 $Y+Z$ 、 $W-X$ 及 $Y-Z$ 的结果。最后的系数可分成四个子频带 (Sub-Bands), LL、HL、LH 及 HH, 如图 2.17 所示, 图中 LL 为低频带, 而 HH 为高频带, 此即为一级 DWT 之后的系数值。跟 DCT 一样, 低频带系数为重建图像重要的信息, 主要用来描绘图像的轮廓, 而高频带系数则相对不重要, 主要用来修饰图像。若要继续进行二级 DWT, 则针对 LL 频带再做水平及垂直方向的处理。JPEG2000 就是将 DWT 变换作为其技术, 选择可分离的滤波器组, 对图像进行三级小波分解, 产生 LH, HL, HH 三个高频带系列, 一个 LL_3 低频带。图 2.18 显示了图像经过三级 DWT 之后, 所形成的 10 个子频带。其中, LL_3 是低频子带, 是最重要的频带, 它表示由小波变换分解级数决定的最大尺度、最小分辨率下对原始图像的最佳逼近, 它的统计特征与原始图像相似, 大部分能量集中在此。高频带则分别是图像在不同尺度、不同分辨率下的细节信息。分辨率越低, 其中有用信息比重越大。对于同一小波级, 低频子频带图像 LL 最重要, 其次是 HL 与 LH, HH 相对最不重要。对于不同小波级, 级高者重要, 级低者不重要。三级小波子频带图像的重要性为 $LL_3 > HL_3 > LH_3 > HH_3 > HL_2 > LH_2 > HH_2 > HL_1 > LH_1 > HH_1$ 。图 2.19 给出了 Lena 图像的一级小波变换和二级小波变换示意图。DWT 跟 DCT 一样, 也是可逆的, 其逆变换的运算如图 2.20 及图 2.21 所示。先将 DWT 系数做垂直恢复, 依图 2.20 的方式将 W 、 X 、 Y 及 Z 四个值, 做相加除以 2 及相减除以 2 的运算, 如此即可恢复到图像做完一阶 DWT 水平变换的系数值。接着再将图 2.20 的结果做水平恢复, 依图 2.21 的方式, 将 A 、 B 、 C 及 D 四个值, 做相加除以 2 及相减除以 2 的运算, 即可恢复原始的图像。若图像进行了三级 DWT 运算, 则需针对每一级的 DWT 运算结果作恢复。

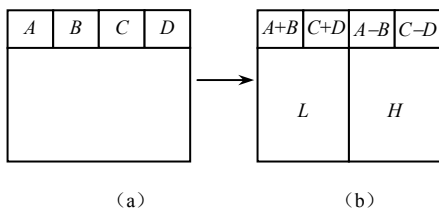


图 2.15 水平方向处理后的 DWT 系数

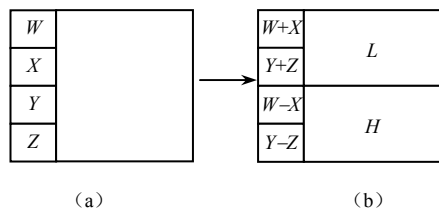


图 2.16 垂直方向处理后的 DWT 系数

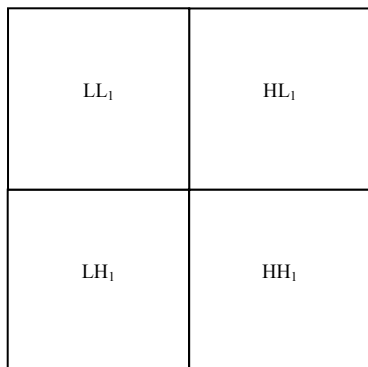


图 2.17 一级 DWT 处理后的子频带

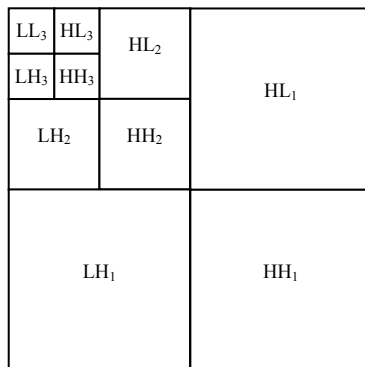
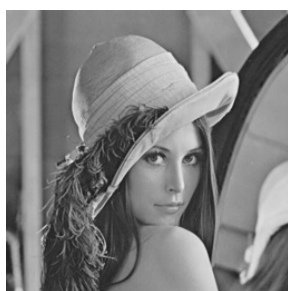
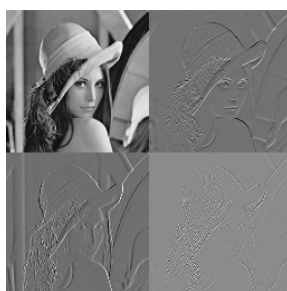


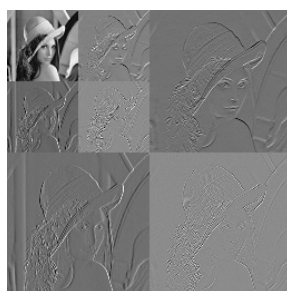
图 2.18 三级 DWT 处理后的子频带



(a) 原始 Lena 图像



(b) 一级小波分解



(c) 二级小波分解

图 2.19 原始 Lena 图像和小波分解图像

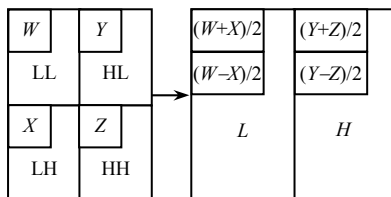


图 2.20 垂直恢复后的 DWT 系数

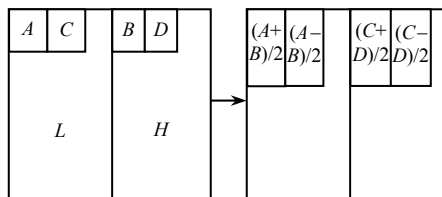


图 2.21 水平恢复后的 DWT 系数

图像 DWT 域隐写术的嵌入算法的基本思路如下：先通过多分辨率分析的小波分解，将原始图像分解到对数间隔的子频带之中，然后对原始图像在每个分辨率等级上进行分割，形成互不相交的像块，再对各像块按照对视觉效果影响的程度嵌入信息，最后对嵌入信息后的小波域图像进行小波逆变换。以一级小波分解为例，秘密信息嵌入的过程可以描述如下。

(1) 对原始图像进行一级小波分解，得到一个低频子图像 LL 和三个高频子图像 HL、LH 和 HH。

(2) 将待嵌入的秘密信息按比特读取。

(3) 将待嵌入秘密信息的比特嵌入到分解后的小波系数高频带的低位中，如果嵌入的长度大于高频部分的长度，则将其剩余部分嵌入到次高频，乃至次次高频。

(4) 对修改后的小波系数进行小波逆变换，便得到隐写图像。

信息提取的过程如下所述。

(1) 对隐写图像进行一级小波分解。

(2) 取分解后的小波系数的高频部分、次高频、次次高频系数的低位，连接起来即

可得到隐藏的秘密信息。

DWT 相对于 DCT 和 DFT, 有很多独特的优点如下。

(1) 良好的时间频率局部性: 图像信号的局部性, 如局部纹理、亮度等, 对于图像分析和处理非常关键, 全图 DCT 变换的局部性很差, 以分块作为变通, 则会导致马赛克效应, 对结果有不良影响, 而小波变换可以保留这些局部特征。

(2) 多尺度变换: 二维小波变换将原图分解为 LL、LH、HL、HH 四部分。对 LL 分量继续分解, 得到多分辨率分解。在图像压缩上, 这样的分解有利于量化后提高压缩比。在信息隐藏中, 秘密信息的嵌入系数选择有多样性。

(3) 计算复杂度: 从全局变换来看, 图像大小 $N \times N$ 较大时, 离散小波变换 DWT 的复杂度为 $O(N)$, DCT 为 $O(N \times \log N)$, DWT 要优于 DCT, 但考虑到有些方法用到了分块方法, 在分块情况下 (N 较小), DWT 的计算开销则比 DCT 大得多。

2.5.4 JPEG 图像隐写术

除了空域隐写术和变换域隐写术, 还有一类隐写术叫做压缩域隐写术。压缩域隐写术是指秘密信息的嵌入、检测和提取都直接在压缩域数据中进行的隐写算法, 比如直接针对 LZW 压缩图像、JPEG 压缩图像、JPEG2000 压缩图像、**矢量量化** (Vector Quantization, VQ) 压缩图像、**块截断编码** (Block Truncation Coding, BTC) 压缩图像。例如, 一个典型的压缩域算法是对 GIF 图像的 LZW 压缩数据直接修改 LSB。由于压缩域算法是对压缩数据的修改, 因此很容易在反压缩时产生失真, 一般情况下使用压缩域进行信息隐藏的软件相对较少。限于篇幅, 这里只讨论针对 JPEG 图像的隐写术。下面先简单介绍 JPEG 压缩的基本原理, 然后介绍针对 JPEG 压缩图像的几种典型隐写术。

1. JPEG 压缩原理

JPEG (Joint Photographic Experts Group) 是 ISO / IEC 和 ITU-T 联合图像专家小组为静态图像建立的第一个国际数字图像压缩标准, 也是至今应用最广的图像压缩标准。JPEG 的压缩模式有以下四种: ① **顺序式编码** (Sequential Encoding), 一次将图像由左到右、由上到下顺序处理; ② **递增式编码** (Progressive Encoding), 可将图像分数次处理, 由粗到细进行编码, 以从模糊到清晰的方式来传送图像; ③ **无失真编码** (Lossless Encoding), 基于 DPCM (差分脉码调制) 进行压缩, 可以保证无失真地重建原始图像; ④ **层次编码** (Hierarchical Encoding), 以各种分辨率对图像进行压缩, 可以根据不同的要求, 获得不同分辨率的图像。基于以上四种模式, JPEG 标准包括两种基本的压缩算法, 一种是基于 DCT 的有损压缩算法, 另一种是基于预测方法的无损压缩算法。由于前者具有较高的压缩率, 已经成为目前 JPEG 压缩的常见形式。

以一幅 24 位彩色图像为例, JPEG 的压缩步骤主要分为: ① DCT 变换; ② 量化; ③ 编码。JPEG 的编码流程如图 2.22 所示。解压缩时, 每一步本质上都是完成编码时对应的逆过程。由于 JPEG 只支持灰度图像和 YUV 颜色模式的数据结构, 所以在将彩色图像进行压缩之前, 必须先对颜色模式进行数据变换, 转换公式如下

$$\begin{cases} Y = 0.299R + 0.587G + 0.114B \\ U = -0.147R - 0.289G + 0.436B \\ V = 0.615R - 0.515G - 0.100B \end{cases} \quad (2.27)$$

其中, R 、 G 、 B 分别为红、绿和蓝分量, Y 为亮度分量, U 、 V 表示两个色度分量。

变换完成之后还需要进行数据采样，一般采用的采样比例是 4:1:1 或 4:2:2。采样过程之后，图像数据量将得到很大程度的压缩。

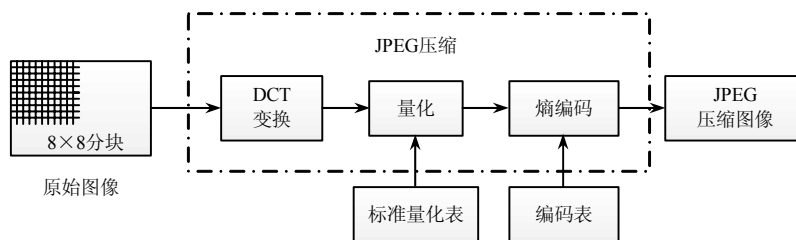


图 2.22 JPEG 基于 DCT 变换的编码流程图

在编码过程中，JPEG 将每个分量图像分成 8×8 的像素块，对各块分别进行 DCT 变换、量化和熵编码。主要步骤如下。

(1) 将原始图像分割为不重叠的 8×8 的小块。

(2) 对每个小块作二维正向 DCT 变换，得到 64 个 DCT 系数，分别代表图像块的不同频率成分，其中左上角为直流（DC）系数，其余 63 个为交流（AC）系数。从左到右、从上到下频率逐渐变大。

(3) 对 DCT 系数进行量化。量化过程中，通常针对不同的频率成分采用不同的量化步长。量化步长的大小决定了原始数据的失真度大小。由于人类视觉对低频较为敏感，所以量化步长由低频到高频呈上升趋势，式 (2.28) 是推荐的标准量化表。

(4) 利用如图 2.14 (a) 所示的之字形扫描方式将 8×8 系数矩阵变成一维数列，频率按由低到高排列。

(5) 使用差分脉码调制（Differential Pulse Code Modulation, DPCM）对直流系数进行编码。

(6) 使用行程长度编码（Run-Length Encoding, RLE）对交流系数进行编码。

(7) 熵编码（Entropy Encoding）。使用熵编码对 DPCM 编码后的直流 DC 系数和 RLE 编码后的交流 AC 系数作进一步的压缩。

$$Q = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix} \quad (2.28)$$

在 JPEG 编码过程中，DCT 变换和量化都是有损变换，而熵编码是无损变换。如果在量化前嵌入隐密信息，经过变换后可能会引起信息丢失，导致解码时不能正确获得秘密信息。因此，目前主要的 JPEG 图像隐写算法的基本原理都是将秘密信息与量化后的 DCT 系数的 LSB 位联系起来，从而达到嵌入和提取秘密信息的目的。针对 JPEG 图像隐写算法的具体步骤为：① 对 JPEG 图像的压缩数据采用哈夫曼解码或算术解码，得到图像的 DCT 量化系数；② 按照某种嵌入规则对 DCT 量化系数做微小的修改，将秘密信息嵌入到量化系数中；③ 对修改后的量化系数表进行熵编码，重新生成压缩数据，即隐写 JPEG 图像。

其中第②步中 DCT 量化系数的修改方法是隐写算法的关键。基于 JPEG 图像的诸多特点，目前出现了很多以它为载体的隐写算法，典型的有 JSteg、F5、OutGuess 和 MB 算法，它们都是依据一定的策略将秘密信息嵌入量化后的系数 DCT 系数上，分别介绍如下。

2. JSteg 算法

JSteg (<http://zooid.org/~paul/crypto/jsteg/>) 是最早以 JPEG 图像为载体的隐写算法，它由 Upham 提出，它利用了 LSB 替换思想，主要思路是：将秘密消息嵌入到量化后的 DCT 系数的最不重要位 (LSB) 上，但对原始值为 -1, 0 和 1 的量化 DCT 系数不进行嵌入，如图 2.23 所示。提取隐密消息时，只需将隐写图像中不等于 -1, 0 和 1 的量化 DCT 系数的 LSB 位提取即可。

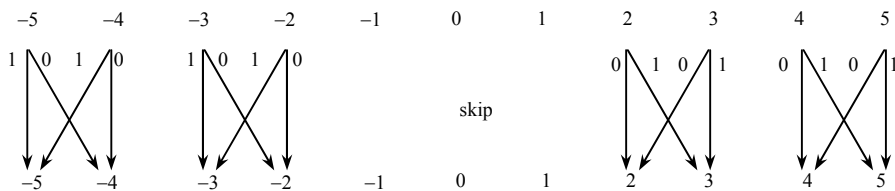


图 2.23 JSteg 嵌入机制

JSteg 算法根据秘密信息嵌入顺序的不同，可分为顺序嵌入法和随机嵌入法。顺序嵌入算法简单，易实现，但安全性很差，因为秘密信息是连续替换载体图像的有效量化 DCT 系数的 LSB，使得载体图像中被替换的部分和未被替换的部分存在统计特性差异，秘密信息的安全性得不到保证，利用 χ^2 分析很容易检测出。而随机嵌入算法应用一个伪随机序列来确定秘密信息的嵌入位置，这样使得秘密信息随机地分布在整个载体图像中，从而提高了秘密信息的安全性。由于 JSteg 算法只是修改绝对值非 0 且非 1 的 DCT 系数值来实现嵌入，所以嵌入容量较小。

3. F5 算法

JPEG 载体图像的量化 DCT 系数通常满足以下三个特性 (图 2.24 给出了直方图示意图): ① DCT 系数的绝对值越大，其对应直方图中的值出现频率就越小; ② 随着 DCT 系数绝对值的增大，其出现频率下降的幅度减小; ③ 各系数出现频率关于 0 对称。

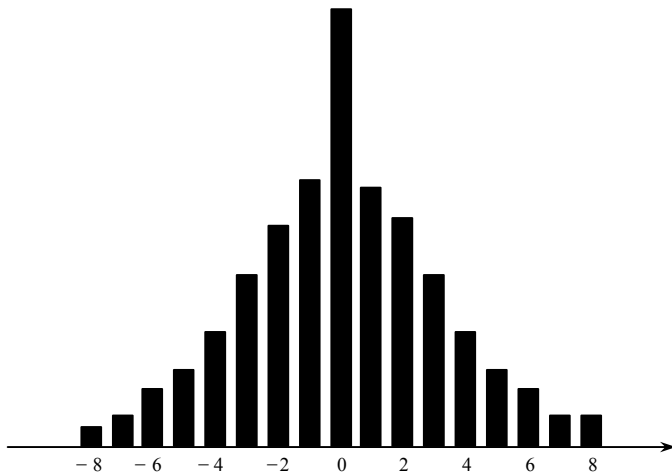


图 2.24 JPEG 载体图像的量化 DCT 系数直方图示意图

F5 隐写算法是由 F3 和 F4 隐写算法发展来的，它在隐写后仍能满足上述 DCT 系数直方图特性要求，故可以抵抗 χ^2 分析。下面依次介绍 F3、F4 和 F5 算法。

(1) F3 算法

F3 算法以 JSteg 为基础进行改进，采用了新的嵌入机制（图 2.25），具体如下。① 载体图像中每个非 0 的 DCT 系数的 LSB，都用来隐藏 1 比特的秘密信息，若秘密信息比特位与 DCT 系数的 LSB 相同，则不进行修改；若不同，则将相应 DCT 系数的绝对值减 1，符号不变。注意：值为 0 的 DCT 系数不用来负载秘密信息。② 若载体图像原 DCT 系数为 +1 或 -1，而待嵌入秘密比特位为 0，那么按①进行修改后，此时原系数值变为 0，则将本次嵌入操作视为无效，重新选择嵌入位。③ 提取时，将隐写图像中不为 0 的 DCT 系数的 LSB 按序取出即为秘密信息。

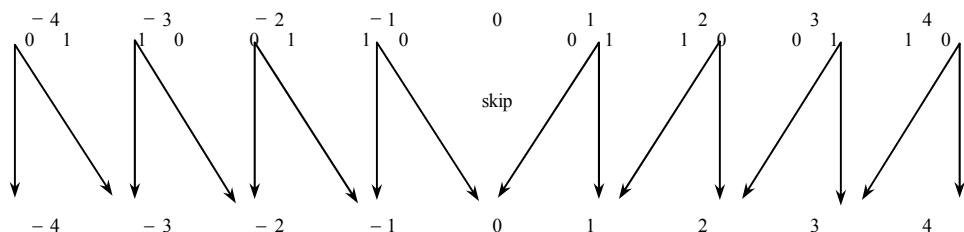


图 2.25 F3 嵌入机制

由 F3 算法的嵌入机制②可知，由于无效隐藏，隐写图像中会嵌入比原秘密信息更多的 0，因此得到的隐写图像 DCT 系数直方图中 $2i$ 位置上的条形柱要比 $2i-1$ 位置上的条形柱高一些，如图 2.26 所示，这很容易引起隐写分析者的怀疑。

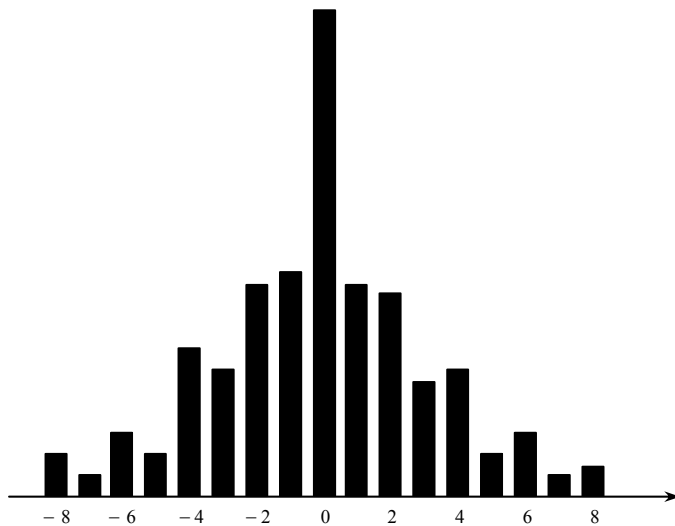


图 2.26 F3 隐写算法作用后图像的 DCT 系数直方图示意图

(2) F4 算法

基于 F3 算法存在的缺陷，**F4 算法**在嵌入策略上作了稍稍的改动，如图 2.27 所示，其区别在于：在 F3 算法中，奇数代表秘密消息 1，偶数代表秘密消息 0；而在 F4 算法中，用正奇数和负偶数代表秘密消息 1，负奇数和正偶数代表秘密消息 0。经改进后，不论嵌入的秘密比特位为 0 还是 1，都可能产生无效隐藏而需重新嵌入，这样自然图像所具

有的 DCT 系数直方图特性就不会被破坏, 如图 2.28 所示。可见 F4 算法的安全性更高。

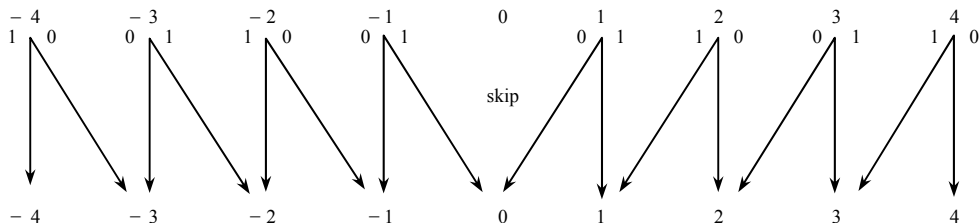


图 2.27 F4 嵌入机制

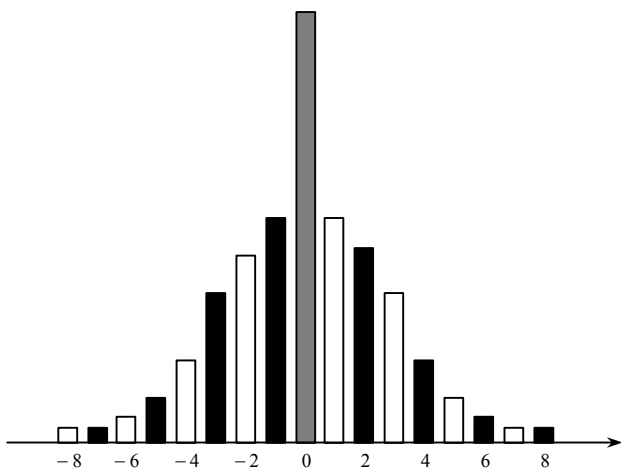


图 2.28 F4 隐写算法作用后图像的 DCT 系数直方图示意图

(3) F5 算法

Westfeld 在 2001 年将 F4 算法作了进一步的改进, 提出了 **F5 算法**^[25]。F5 算法整合了混洗技术和矩阵编码技术, 使得算法安全性更高。矩阵编码方式(1, n , l)是将 l 比特秘密信息嵌入到 $n=2^l-1$ 个符合要求的 DCT 系数中, 最多只需修改一个 LSB 的编码方式。以 $l=2$ 时的简单情况为例, 设 b_1 、 b_2 是待嵌入的两个秘密比特位, a_1 、 a_2 、 a_3 分别是三个 DCT 系数的 LSB, “ \oplus ”表示异或运算符, 则有下列运算规则: ① 如果 $b_1=a_1 \oplus a_3$, $b_2=a_2 \oplus a_3$, 则不改变原始数据; ② 如果 $b_1 \neq a_1 \oplus a_3$, $b_2=a_2 \oplus a_3$, 则改变 a_1 ; ③ 如果 $b_1=a_1 \oplus a_3$, $b_2 \neq a_2 \oplus a_3$, 则改变 a_2 ; ④ 如果 $b_1 \neq a_1 \oplus a_3$, $b_2 \neq a_2 \oplus a_3$, 则改变 a_3 。提取秘密信息时, 只需进行逆操作, 即 $b_1=a_1 \oplus a_3$, $b_2=a_2 \oplus a_3$ 。这里用 R 表示载体数据利用率, 即嵌入的秘密信息比特数与使用的载体数据的 LSB 数的比值, 则有

$$R = \frac{l}{2^l - 1} \quad (2.29)$$

用 E 表示嵌入效率, 即每修改一个 LSB 可以平均嵌入的秘密比特数, 则有

$$E = \frac{2^l}{2^l - 1} l \quad (2.30)$$

注意到, 嵌入 2 比特的秘密信息平均只需修改 $3/4$ 个 LSB, 而普通的 LSB 隐写需要修改一个 LSB, 由式 (2.30) 可求得嵌入效率提高了。而 F5 算法应用矩阵编码, 目的就是为了提高 LSB 隐写算法的嵌入效率, 但有一个缺陷就是载体的利用率降低了。表 2.5 给出了不同 $(1, n, l)$ 对应的矩阵编码的性能, 很明显, 嵌入效率和载体数据利用率成反比。

下面给出 F5 算法的具体执行过程。

1) 对 JPEG 图像进行部分解码, 得到量化后的 DCT 系数。

2) 利用密码技术, 对 DCT 系数进行混洗。

3) 计算确定矩阵编码参数 l 、 n , 其中 $n=2^l-1$ 。

4) 利用矩阵编码技术嵌入秘密信息如下。

4.1) 取出待嵌入的 l 比特秘密信息, 使用伪随机数生成器生成的伪随机序列与 l 比特秘密信息进行异或运算, 得到 l 比特随机化 $\{0, 1\}$ 流, 并取出 n 个非 0 的 DCT 系数。

4.2) 根据矩阵编码计算是否需要修改 DCT 系数, 如果不需要, 则返回 4.1) 进行下一组嵌入; 如果需要, 则将要修改的 DCT 系数的绝对值减 1, 符号不变。

4.3) 判断 4.2) 中被修改的 DCT 系数是否变为 0, 若没有, 则返回 4.1) 进行下一组嵌入; 若系数变为 0, 则视本次隐藏无效, 要重新选择一个非 0 的 DCT 系数, 与其余 $n-1$ 个非 0 的 DCT 系数组成 n 个 DCT 系数, 返回步骤 4.2) 继续嵌入。循环执行 4.1)、4.2) 和 4.3), 直至嵌入完成。注意: 在嵌入过程中, 不考虑 DC 系数以及值为 0 的 AC 系数。

5) 嵌入完成后, 对 DCT 系数进行逆混洗、编码, 产生隐写后的 JPEG 图像。

相对而言, F5 算法是一种较安全可靠的隐写算法, 但是也同样存在缺陷。F5 算法与 F4 算法类似, 都采用了 DCT 系数绝对值减 1 的方法来隐藏秘密信息, 这样会产生更多的 DCT 系数为 0 的值, 出现收缩效应, 如图 2.29 所示。

表 2.5 不同的 (l, n, l) 对应的矩阵编码的性能

l	1	2	3	4	5	6	7	8	9
n	1	3	7	15	31	63	127	255	511
$R(\%)$	100.00	66.67	42.86	26.67	16.13	9.52	5.51	3.14	1.76
E	2.00	2.67	3.43	4.27	5.16	6.09	7.06	8.03	9.02

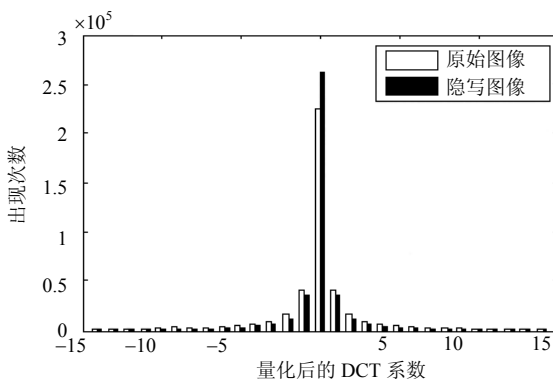


图 2.29 F5 隐写算法作用后某载体图像的 DCT 系数直方图变化

4. Outguess 算法

基于上述隐写算法存在的缺陷, 研究者提出了一类安全的 JPEG 隐写算法, 它们主要考虑如何在嵌入秘密信息后, 最大限度地保持载体图像的概率统计分布。OutGuess 算法就是其中一种典型的方法。

Outguess (<http://www.outguess.org>) 是由 Neils Provos 提出的一种针对 JSteg 类算法缺陷的隐写算法, 该算法的嵌入过程是在 DCT 变换和量化之后, 哈夫曼编码之前进行秘

密信息嵌入。嵌入大体分两步：第一步是嵌入。将秘密信息嵌入到载体图像量化后的 DCT 系数中，嵌入时只是将秘密信息比特位与随机选择的非 0、1 的 DCT 系数的 LSB 进行简单替换。第二步是纠正。嵌入完成后，对载体图像未修改 DCT 系数部分进行纠错编码以修改嵌入过程带来的载体 DCT 系数直方图特性的改变，使得隐写图像的 DCT 系数直方图与原载体图像的 DCT 系数直方图一致。OutGuess 算法在一定意义上保持了直方图特性，能够抵抗 χ^2 分析，算法安全性得到提高，但是由于嵌入和纠正两步均修改了 DCT 系数，所以会引起图像空间域的边界不连续性，且嵌入信息越多带来的变化越大，这同样会引起隐写分析者的怀疑。

5. MB 算法

基于模型的隐写术 (Model Based Steganography) 是由 Sallee 提出^[26]的一种通用的隐写算法，又称 MB 隐写术。MB 利用载体图像的统计模型来嵌入隐密信息，它对抗一阶统计攻击。MB 算法假设 JPEG 图像所有 8×8 系数块的同一位置非零 AC 系数可以用一个有两个参数 p 和 s 的广义柯西分布来拟合

$$P(u) = \frac{p-1}{2s} \left(\left| \frac{u}{s} \right| + 1 \right)^{-p} \quad (2.31)$$

其中， u 表示非零 AC 系数值， p 和 s 为参数， $p > 1$ ， $s > 0$ 。对应的累积分布函数如下：

$$D(u) = \begin{cases} 0.5(1 + |u/s|)^{1-p} & u \leq 0 \\ 1 - 0.5(1 + |u/s|)^{1-p} & u > 0 \end{cases} \quad (2.32)$$

对于给定的载体 JPEG 图像和秘密信息，假设嵌入到所有 8×8 系数块的某个相同位置的非零 AC 系数，则 MB 隐写嵌入算法步骤可以描述如下（当然，可以选择多个位置嵌入，只需重复执行以下步骤）。

(1) 求 AC 系数的低精度直方图。这里，低精度直方图的柱宽 (Bin Size) 大于 1，并称之为嵌入阶距。注意，值为 0 的柱宽认为是 1，因为值为 0 的 AC 系数直接跳过不嵌。这样，每个 AC 系数都可以用柱索引 (Bin Index) 和柱内偏移量 (Offset) 表示。所有柱索引组成 X_α ，这部分信息在隐写过程中保持不变。

(2) 基于广义柯西分布利用最大似然法来匹配 X_α ，得到式 (2.31) 的参数 p 和 s 。

(3) 为每个系数在它对应的低精度直方图柱体内的偏移量指定一个符号，这些符号组成 X_β 。若嵌入阶距为 2，则 X_β 中的符号取值只有两种可能，不是 0 就是 1。使用累积分布函数式 (2.32) 来计算 X_β 中每个系数的各种可能符号的概率值。

(4) 使用伪随机数生成器确定修改系数的块顺序。这一步是为了增加安全性。

(5) 把秘密信息和步骤 (3) 中计算得到的符号概率一起按照步骤 (4) 得到的修改顺序输入到非自适应的算术解码器中，得到每个系数在低精度直方图柱体内的新偏移量，记得到的符号为 X_β' 。这里，采用了非自适应的算术编码的解码操作。所谓非自适应是指与经典的算术编码^[27]中的符号概率按照自适应估计得到的不一样，这里是直接赋予的。

(6) 基于 X_α 和 X_β' 计算所有的新系数值。

对于给定的可疑伪装 JPEG 图像，MB 隐写术的秘密信息抽取过程的前 4 步与嵌入过程一样，然后把前面计算得到的符号和符号概率按照修改顺序输入到一个非自适应的算术编码器中，得到秘密信息。MB 隐写术可以很好地保存载体图像的一维统计特征，在某些情况下比 F5 隐写术更为优越，如图 2.30 所示。MB 隐写前后 DCT 系数直方图统计几乎没发生变化，隐藏信息后能够保持原始载体图像直方图统计特性，因此 MB 隐写能

够抵抗卡方分析、单直方图统计等常规隐写分析方法。MB 在 DCT 系数上进行隐写，单直方图统计未发生明显变化，但隐写信息时，DCT 系数进行了修改，DCT 块内相邻系数相关性发生了变化，可通过统计相邻 DCT 系数相关性来进行隐写分析。另外，与载体图像相比，MB 隐写图像和统计模型更加吻合，因此，通过分析待检测图像与统计模型的接近程度，也可实现对 MB 算法的检测。

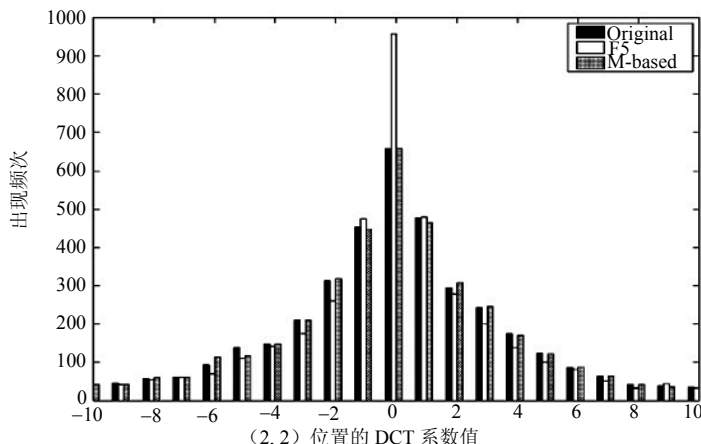


图 2.30 F5 隐写算法和 MB 算法作用后某 JPEG 图像的(2,2)位置 DCT 系数直方图比较

2.6 基于音频载体的隐写术

2.6.1 引言

音频文件相对于其他载体具有其自身独特的特点和要求。一方面，人耳听觉系统 (Human Audio System, HAS) 对音频文件中的加性随机噪声非常敏感，能够觉察出微小的扰动，也就是说音频文件作为载体信号冗余度很低，即便在时域对音频信号作微小的改变，也会对其清晰度造成比较大的影响，因此对音频文件进行信息嵌入在实现透明性上难度较大，必须充分利用人耳听觉系统特性。另一方面，人耳听觉系统存在掩蔽特性，即当两个声音同时存在，或者在相隔很短的时间内存在时，相对微弱的声音信号会被幅度更强的声音信号所掩蔽的一种听觉现象。同样，在频域中，如果两个信号的频率相近，较弱的信号也会因为相邻近的较强信号的存在而变得不可察觉。人耳的这些听觉特性以及现代音频编码技术的不完善使得音频编码中存在着一定的冗余，同时也为在音频中进行信息隐藏提供客观的条件。当需要在音频文件中嵌入秘密数据时，我们可以利用人耳的听觉掩蔽曲线，使隐藏进去的秘密数据的幅度保持在人耳所不能察觉的范围之内。

下面简要介绍与隐写相关的人耳听觉系统特性，包括人耳听觉阈值、听觉时域掩蔽效应和频域掩蔽效应。首先，需要介绍声压级 (Sound Pressure Level, SPL) 的概念。声压级用来表示声音的能量大小，它取决于大气中的声音与参考声压 $p_0=20\mu\text{Pa}$ 的比值

$$\text{SPL} = 20 \log \frac{p}{p_0} \quad (2.33)$$

人耳存在听觉绝对阈值，它用声压级为单位，如图 2.31 所示，可用如下非线性函数来逼近

$$q(f) = 3.64f^{-0.8} - 6.5e^{-0.6(f-3.3)^2} + 10^{-3}f^4 \quad (2.34)$$

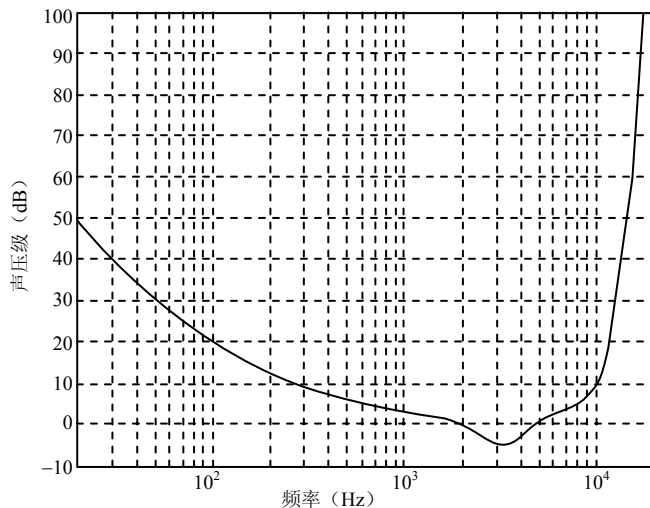


图 2.31 听觉系统在安静环境下的绝对听觉阈值

正常人可感知的声音频率范围为 20Hz~20kHz，人耳对声音内容的感知则由声音所含频率成分决定，人耳可完成对声音的频谱分析。人耳耳蜗的机械构造和动作特性构成了一个非等宽的机械滤波器组，每个滤波器都相当于一个子带滤波器。频域掩蔽又称为同时掩蔽，如图 2.32 所示，其机理如下：耳蜗基膜上的某个临界带对一个强信号产生了足够大的响应，使得该临界带和相邻临界带对其他较弱信号的响应能力降低。实际中，有三种常见的频域掩蔽效应：纯音掩蔽噪声、噪声掩蔽纯音、相邻频带间的掩蔽效应。与频域掩蔽相对应的是时域掩蔽，也可称为非同时掩蔽。时域掩蔽的机理如图 2.33 所示：一个掩蔽音能减弱人耳对其前一段时间和其后一段时间的声音的响应。时域掩蔽的有效区域包括在掩蔽音出现前 1~2ms 的前掩蔽区域和掩蔽消失后 50~300ms 的后掩蔽区域。

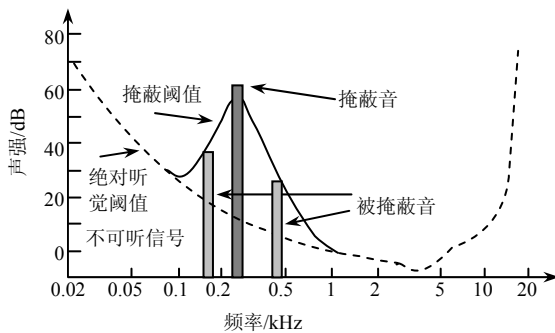


图 2.32 频域掩蔽效应

以音频信号为载体的隐写方法按照嵌入域的不同可分为压缩域和非压缩域的嵌入。目前大部分的研究都集中在非压缩域的嵌入算法上，因为非压缩域的数据冗余和感知冗余较大，可以隐藏较多的信息量，而压缩域正是通过去除这些冗余来达到音频压缩的目的。因此，对于音频，压缩域的嵌入比非压缩域的嵌入要困难。其中，非压缩域的方法又可分为时间域方法和变换域方法。下面按照不同的嵌入域分类介绍常见的音频隐写术。

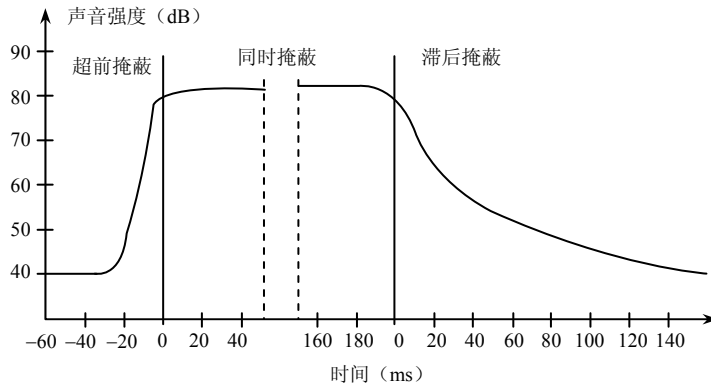


图 2.33 时域掩蔽效应

2.6.2 时域隐写方法

时域音频隐写方法就是在原始音频信号的时间域上嵌入秘密信息，该方法的特点是计算复杂度低，但对音频压缩和滤波等一般信号处理方法的鲁棒性较差。常用的时域隐写方法有：最不重要位（LSB）算法、回声隐藏、扩频算法和量化索引调制算法等。

1. LSB 嵌入法

LSB 方法是将秘密信息嵌入到载体对象中的一种最简单的方法。任何秘密数据都可转换成一串二进制码流，而音频的每个采样数据也是用二进制来表示的。这样，我们可以通过将音频信号部分采样值的 LSB 用秘密信息的二进制位替换，从而达到将秘密信息隐藏到音频数据中去的目的。提取时只需要把采样点相应的 LSB 取出来，就可以提取出秘密信息。

为了加大对秘密信息攻击的难度，还可以引入一个伪随机序列来控制嵌入位置。伪随机信号可以由伪随机序列发生器来产生。当伪随机序列发生器的结构固定时，不同的初始值会产生不同的伪随机序列，这样收发方只需要秘密地传送一个初始值（作为密钥），而不需要传送整个伪随机序列值。只要能保证是合法用户才能得到该密钥，则根据 Kerchoff 法则可知系统是安全的。这样，任何一个企图提取出秘密信息的第三方在不知道密钥的情况下，难以达到攻击的目的。

LSB 方法本身简单易实现，信息嵌入和提取算法简单、速度快。LSB 方法的优点是算法简单、速度快，且在音频信号中可隐藏的数据量大，例如一个 16 位脉冲编码调制（Pulse Code Modulation, PCM）编码精度、采样频率为 44.1kHz 的音频，在每个采样点中嵌入 1 位的信息，就能够达到 44.1Kbps 的嵌入率，但这种方法的最大缺陷是对信号处理的鲁棒性很差，信道干扰、数据压缩、滤波、重采样等操作都可能破坏秘密信息。在实用中，LSB 方法只能用在封闭的数字到数字的环境下。为提高鲁棒性，可以将秘密信息嵌入到音频信号的较高位，但这样带来的结果是大大降低了秘密信息的隐蔽性，即不可听性。为了改善这一点，可以在嵌入过程中根据音频信号的能量进行数据嵌入位的选择，这种方法对平均能量较高的音频文件较为有效。另外，LSB 方法的安全性较弱，经过 LSB 嵌入后的音频在某些方面会呈现出一定的规律性，如直方图的变化，这样就为隐写分析提供了依据。

2. 回声隐藏法

回声隐藏（Echo Hiding）是利用时域掩蔽效应来进行隐写的典型算法。回声隐藏的基本思想是通过引入回声将秘密数据嵌入到载体音频文件中。它利用了人耳听觉系统在

时域上的滞后掩蔽特性，即一个音频信号中，如果弱信号在强信号后很短的时间内（一般 0~200ms）出现，弱信号会变得不可听见。回声隐藏正是利用人耳的这一听觉特性，试图在离散时间音频信号中引入回声，将秘密数据隐藏在回声中。通过回声的数量、位置和幅度大小的不同来表示秘密信息，实现信息隐藏的目的。嵌入公式可以描述如下

$$s(n) = c(n) + \alpha \cdot c(n - \delta) \quad (2.35)$$

其中， $c(n)$ 为原始音频信号， $s(n)$ 为隐写音频信号， α 是回声相对于原始音频信号的衰减系数， δ 为延时参数，一般表示回声滞后于原始音频信号的时间间隔。采用不同的延时参数分别代表嵌入“0”或“1”。原始音频信号 $c(n)$ 和经过回声隐藏的隐写信号 $s(n)$ 相对于人耳来说，前者就像是从小耳机里听到的声音，没有回声，而后者就像是从小扬声器里听到的声音，由所处空间诸如墙壁、家具等物体产生的回声。由此可见，回声隐藏与其他方法不同，它不是将秘密数据当作随机噪声嵌入到载体数据中的，而是利用载体数据的环境特征（回声）来嵌入秘密信息。回声隐藏算法的秘密信息嵌入流程如图 2.34 所示。

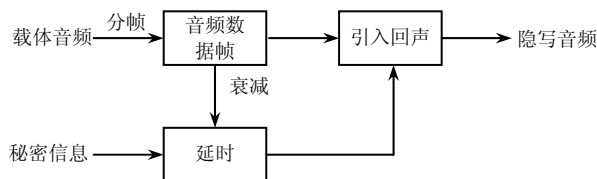


图 2.34 回声隐藏算法嵌入流程图

回声提取的关键是回声延时的测定，不同的延时参数分别代表“0”或“1”，因此根据回声延时来得到嵌入的秘密信息。由于引入回声后的音频信号可以看成原始音频信号与回声核的卷积，所以常用倒谱自相关的方法来测定回声间距。

回声隐藏将秘密信息作为载体信号的回声，对一些有损压缩算法具有一定的鲁棒性。但普通的回声算法存在一定的缺陷，一方面如果原始音频信号中本来就含有回声，提取时对于回声延时的测定可能不准确，容易出现误码；另一方面当回声幅度较小时，采用传统的倒谱分析检测法，回声在变换域的尖峰容易被淹没，因此很难检测到，而如果增大回声幅度，隐藏效果又会降低。此外，回声隐藏较之 LSB 方法，其隐写容量相对较低。

3. 扩频法

扩频法的基本思想是把窄带的秘密数据扩展到载体信号的整个频率谱上，其中常用的是直接序列扩频（DSSS）技术。该技术使用一个密钥产生一个伪随机序列（PN 序列），理想的 PN 序列是功率谱密度在整个频谱范围内均匀分布的白噪声，用它来调制秘密信息序列，得到扩展频谱序列，在提取时也使用该密钥解码。经过载波和伪随机序列调制的秘密信息，乘以一定的衰减系数，一般衰减至原始音频信号幅度的 0.005 左右，再叠加到每一帧的载体音频信号上去，提取采用相关的方法。基于 DSSS 方法的整个流程如图 2.35 所示。

下面介绍一下 DSSS 方法的过程。设长度为 N 的秘密信息序列 $\{m(1), m(2), \dots, m(N)\}$, $m(i) \in \{-1, 1\}$, $1 \leq i \leq N$ 为双极性序列。设扩频的片率（Chip Rate）为 L ，也就是说每一位秘密信息重复 L 次，这样秘密信息序列扩展后的长度为 $L \times N$ ，记作 $m(n)$, $1 \leq n \leq L \times N$ 。伪随机序列 $R(n)$ 由一个伪随机数发生器产生，长度也为 $L \times N$ 。发生器的种子由一组密码构成，且假定密码只有授权方才知。秘密信息在嵌入前经扩频序列 $r(n)$ 调制成如下的形式

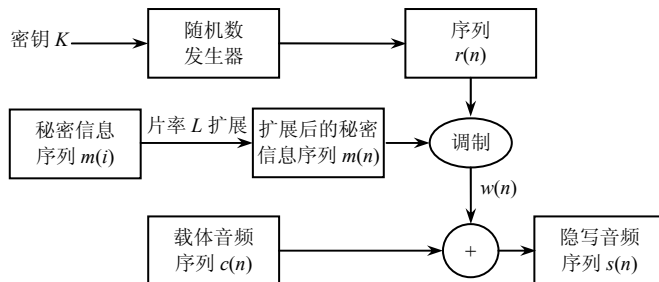


图 2.35 DSSS 隐写方法流程

$$w(n) = \alpha \cdot m(n) \cdot r(n), \quad 1 \leq n \leq L \cdot N \quad (2.36)$$

其中, α 为嵌入强度, $w(n)$ 为调制后的待嵌入序列。经过调制后的序列 $w(n)$ 可以直接应用加法公式嵌入载体音频序列 $c(n)$ 上, 得到隐写音频序列 $s(n)$ 。

$$s(n) = c(n) + w(n), \quad 1 \leq n \leq L \cdot N \quad (2.37)$$

在提取秘密信息时, 假定接收方知道秘密信息的嵌入位置及调制序列 $r(n)$ 产生时的密码种子, 这样就可以重新生成序列 $r(n)$ 。用收到的可能含有秘密信息的音频载体 $s'(n)$ 与 $r(n)$ 做相关运算来检测秘密信息, 按片率 L 分段计算相关系数 ρ_i , $1 \leq i \leq N$ 如下

$$\rho_i = \sum_{n=(i-1) \cdot L+1}^{i \cdot L} \alpha s'(n) r(n), \quad 1 \leq i \leq N \quad (2.38)$$

并做如下判决得到恢复的秘密信息

$$m'(i) = \begin{cases} 1 & \rho_i > T \\ -1 & \text{其他} \end{cases}, \quad 1 \leq i \leq N \quad (2.39)$$

其中, T 为给定的阈值。

由此可见, 扩频技术的运用将待隐藏的秘密信息扩展为宽频的信号分布于整个音频信号的频谱中, 降低了秘密信息的能量密度, 使他人难于察觉, 同时它也提高了信号的抗干扰能力, 而且该方法易于与加密技术结合进一步提高保密性。

4. 量化索引调制

2001 年由 Chen 和 Wornell 提出的**量化索引调制** (Quantization Index Modulation, QIM) 的大致思想是^[28]: 将秘密信息的可能值看成是量化索引, 如秘密信息的二进制位 0、1 对应的量化索引值为 1、2。每一个索引值对应不同的量化器, 对载体信号进行量化。QIM 隐写方法可描述为: 首先将载体音频信号进行分段, 每一段欲隐藏秘密信息的一位, 用欲隐藏的这个位对应索引的量化器量化载体信号, 即得到嵌入秘密信息位的信号, 依次对每段载体音频按照嵌入位做相应的量化即可实现信息隐藏。提取时应知道分段方法, 首先将隐写音频按同样的方法分段, 然后对该段信号按各种可能的索引对应的量化器进行量化, 比较各种量化结果与该段信号的差异, 取差异最小的量化索引值, 其对应的二进制位即判为该段嵌入的秘密信息值。对每一段依次进行提取, 即得到秘密信息序列。

对上述基本的 QIM 方法的改进可以得到**抖动调制** (Dither Modulation, DM) 方法。所谓抖动调制是根据欲隐藏的秘密信息位来调制量化区间, 对于时域音频信号, 一般采用双极性抖动调制的方法。其原理如图 2.36 所示, 其量化过程可以描述如下。

(1) 选取量化步长 Δ 将待量化音频信号分成如图 2.36 所示的“0”和“1”区间集。

(2) 选取量化步长 Δ 对欲量化的音频信号 c 进行除法, 取整数商 n 和余数 γ , 即 $n = \lfloor c/\Delta \rfloor$, $\gamma = c - n\Delta$ 。

(3) 对音频信号进行量化, 根据秘密信息位 m 的取值和待量化的音频信号 c 的值进行量化, 量化结果可以表示为以下 4 种情况。

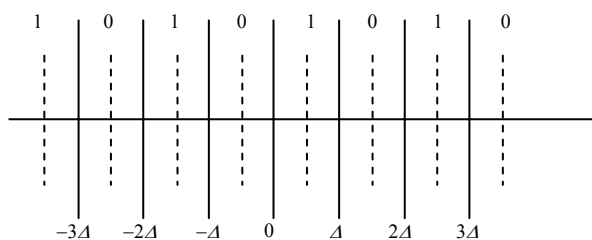


图 2.36 双极性抖动量化调制原理

① 当 $c \geq 0$, $m=1$ 时

$$s = \begin{cases} 2k\Delta + 0.5\Delta, & n = 2k \\ 2k\Delta + 0.5\Delta, & n = 2k + 1 \text{ 且 } |\gamma| \leq 0.5\Delta \\ 2k\Delta + 1.5\Delta, & n = 2k + 1 \text{ 且 } |\gamma| > 0.5\Delta \end{cases} \quad (2.40)$$

② 当 $c \geq 0$, $m=0$ 时

$$s = \begin{cases} 2k\Delta + 1.5\Delta, & n = 2k + 1 \\ 2k\Delta - 0.5\Delta, & n = 2k \text{ 且 } |\gamma| \leq 0.5\Delta \\ 2k\Delta + 1.5\Delta, & n = 2k \text{ 且 } |\gamma| > 0.5\Delta \end{cases} \quad (2.41)$$

③ 当 $c < 0$, $m=1$ 时

$$s = \begin{cases} -2k\Delta - 1.5\Delta, & n = -(2k + 1) \\ -2k\Delta + 0.5\Delta, & n = -2k \text{ 且 } |\gamma| \leq 0.5\Delta \\ -2k\Delta - 1.5\Delta, & n = -2k \text{ 且 } |\gamma| > 0.5\Delta \end{cases} \quad (2.42)$$

④ 当 $c < 0$, $m=0$ 时

$$s = \begin{cases} -2k\Delta - 0.5\Delta, & n = -2k \\ -2k\Delta - 0.5\Delta, & n = -(2k + 1) \text{ 且 } |\gamma| \leq 0.5\Delta \\ -2k\Delta - 1.5\Delta, & n = -(2k + 1) \text{ 且 } |\gamma| > 0.5\Delta \end{cases} \quad (2.43)$$

由此可见, 秘密信息位 m 由量化结果 s 所位于的区间确定, 如果 s 位于 0 区间集, 则表示秘密信息位为 0; 如果 s 位于 1 区间集, 则表示秘密信息位为 1。在量化过程中, 为减少音频信号的失真, 将音频信号量化为与之最接近的 0 区间或 1 区间的中值。

2.6.3 变换域隐写方法

基于变换域的信息隐藏方法在图像隐写术上已经得到了广泛的应用, 现在也越来越多地被应用于以音频文件为载体的保密通信中。变换域隐写方法的基本思想是首先对载体信号进行某种类型的变换运算, 包括离散余弦变换、离散傅里叶变换、离散小波变换等, 通过修改变换域的系数来嵌入秘密信息, 然后进行逆变换, 从而得到隐写后的音频信号。一般来说, 音频低频区的能量较高, 对低频系数的修改可能会影响音频的感知效果, 而选择在音频高频区嵌入秘密信息鲁棒性又较差, 所以通常选择中低频区的系数进行嵌入, 从而嵌入数据既有鲁棒性, 又满足不可感知性。变换域的嵌入算法和提取算法的一般模型如图 2.37 和图 2.38 所示, 秘密信息嵌入的主要步骤如下: ① 应用 DFT、

DCT、DWT 等变换将原始音频信号变换到频域空间；② 选择 N 个频域系数用于隐藏秘密信息；③ 根据一定的规则或公式改变所选择的 N 个频率系数；④ 进行逆变换得到隐写信号。下面简单介绍一下几种典型的变换域隐写方法。

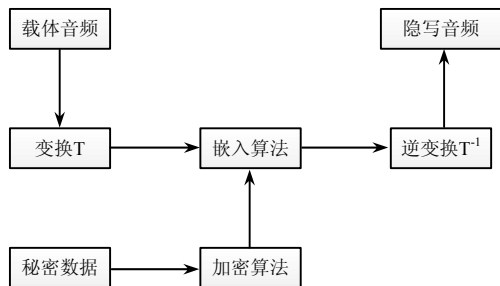


图 2.37 变换域嵌入方法框图

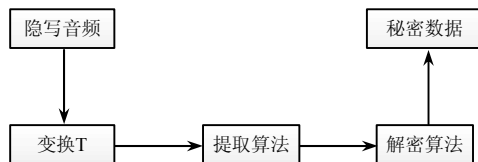


图 2.38 变换域提取方法框图

1. DFT 域隐写方法

该类方法先对音频信号进行 DFT 变换，然后选择中频段（2.4~6.4kHz）的频谱系数来进行秘密数据的嵌入，嵌入过程可直接用表示秘密信息序列的频谱分量来替换相应的频谱系数，也可以采用 DSSS 等方法将扩频调制后的秘密信息加到相应的傅氏变换系数上。这种方法对噪声、录音失真等都具有一定鲁棒性。

2. DCT 域隐写方法

该类方法在音频信号的 DCT 变换系数中有选择性地嵌入秘密数据，例如我们可以建立一种 DCT 噪声信号模型，通过定义 DCT 系数的噪声敏感度，建立秘密信息嵌入位置和嵌入秘密信息后的音频信号的听觉感知性之间的关系，根据隐写的不可感知性的要求选择最优的嵌入位置，然后调节嵌入强度来满足鲁棒性的要求。这种方法能够较好地达到秘密数据的隐蔽性，它对加噪、滤波等处理具有一定的鲁棒性。

3. DWT 域隐写方法

该类方法先对原始音频信号进行多级小波分解，得到不同级别的细节分量和近似分量，然后在细节分量中将经过处理后按一定的规则待隐藏的秘密信息进行嵌入，再通过小波重组，得到隐写音频信号。也可以采用小波包分解的方式，将秘密信息嵌入到小波包高频系数中去。这种方法的鲁棒性良好，能抵抗多种类型的攻击，如加噪、滤波、剪切、重采样、几何变形和有损压缩等。

4. 相位编码方法

人耳听觉系统对声音的相位不太敏感这一事实在大量数字声音压缩系统中被广泛应用。**相位编码**（Phase Coding）方法^[16]通过载体音频相位谱的一个相位转换来代表一个数据，用代表“0”和“1”的参考相位来取代原始音频的绝对相位，以保持各段间的相对相位不变的原则对其他的音频段进行相位的调整，从而达到嵌入秘密数据的目的。相位编码具体步骤如下。

设待秘密信息位为 m ，为了嵌入这一信息位，载体音频信号 $c(n)$ 首先被分成 L 个等长的短序列 $c_i(n)$ ，其中 $c_i(n)$ 长度为 l 。对 $c_i(n)$ 进行 DFT 变换，得到幅度 $A_i(k)$ 和相位 $\varphi_i(k)$

$$F_i(k) = \text{DFT}\{c_i(n)\} = A_i(k) \exp(\varphi_i(k)) \quad (2.44)$$

秘密信息 m 的嵌入需要改变第一个信号片段的相位如下

$$\varphi_1(k) = \begin{cases} \pi/2 & m=0 \\ -\pi/2 & m=1 \end{cases} \quad (2.45)$$

由于两个连续信号片段间的相位变动很容易被检测出来, 因此它们的相位差需要在嵌入秘密信息的音频信号中保持不变。于是, 后续信号片段的相位改变如下

$$\begin{aligned} \varphi'_2(k) &= \varphi'_1(k) + [\varphi_2(k) - \varphi_1(k)] \\ \varphi'_3(k) &= \varphi'_2(k) + [\varphi_3(k) - \varphi_2(k)] \\ &\dots \\ \varphi'_N(k) &= \varphi'_{N-1}(k) + [\varphi_N(k) - \varphi_{N-1}(k)] \end{aligned} \quad (2.46)$$

实际上, 秘密信息的嵌入过程是修改了所有后续信号片段的绝对相位, 以保证它们的相对相位差不变。最后, 新的相位序列 $\varphi'_i(k)$ 和原来的傅里叶幅度序列 $A_i(k)$ 通过 IDFT 来构造隐藏有秘密信息的隐写信号。

$$c'_i(n) = \text{DFT}^{-1}\{F'_i(k)\} = A_i(k) \exp[\varphi'_i(k)] \quad (2.47)$$

提取秘密信息时先对信号进行同步 (即知道信号片段起点和序列长 l), 然后根据 DFT 得到初始相位, 并把该相位和两个参考相位比较以得到 “0” 或 “1”。这种方法能够隐藏的信息非常有限, 因此一般用于音频水印。

相位编码方法具有一定的鲁棒性, 但也存在一个明显缺点, 当代表秘密信息的参考相位急剧变化时, 会出现明显的**相位离差** (Phase Dispersion), 它不仅会影响秘密信息的隐蔽性, 还会增加接收方译码难度。造成相位离差的一个原因是用参考相位代替原始相位而带来变形, 另一个原因是对原始音频信号的相位改动频率太快。为了使相位离差的影响得以改善, 需要在相位值的改变点之间留有一定的间隔, 以使相位的转换变得平缓, 但它的缺点是降低了信息隐写容量, 因此必须在数据嵌入量和嵌入效果之间进行折衷。

2.6.4 压缩域隐写方法

随着信息隐藏研究的不断发展, 使用未压缩音频作为载体的方法逐渐暴露出缺点: 未压缩格式的音频体积较大, 而且传送这种网上并不常见格式的音频本身就会引起怀疑。目前基于感知模型的音频压缩编码方法已被广泛地使用, 能够对音频进行效率较高的压缩, 从而可以减小存储空间或者减小音频传输所用的带宽和时间。音频感知编码技术已经在网络和无线传输中得到了广泛的应用, 目前, 互联网上传播的音频大多是以 MP3 为代表的压缩格式的音频, 因此研究压缩域隐写方法具有十分重要的意义。而许多压缩音频编码本身已经利用听觉特性降低了数据的冗余度, 因此在音频压缩域嵌入的数据量比未压缩域的少, 研究如何在音频压缩域中隐藏信息比在未压缩域中困难得多。在音频压缩域中嵌入秘密数据, 一般有三种方法, 如图 2.39 所示。

嵌入方法一的秘密数据嵌入是在非压缩域上进行的, 即先嵌入后压缩。该方法在压缩过程中容易造成信息丢失, 导致提取误码率的上升。嵌入方法二是直接在压缩音频的码流上嵌入秘密数据, 不需要经过压缩和解压缩过程, 这类方法的嵌入速度较快, 但鲁棒性却不高, 解压缩—重压缩处理会对秘密数据造成破坏。嵌入方法三先将压缩音频解压, 在非压缩域中嵌入秘密数据, 然后重新进行音频压缩, 得到嵌入信息后的压缩音频。这类方法可以提高鲁棒性, 但时间复杂度较高, 因为解压缩和重压缩过程要耗费较多的时间。压缩域音频以 MP3 这种音频格式最为常见, 因此以 MP3 为载体嵌入秘密信息引起人们较多关注。

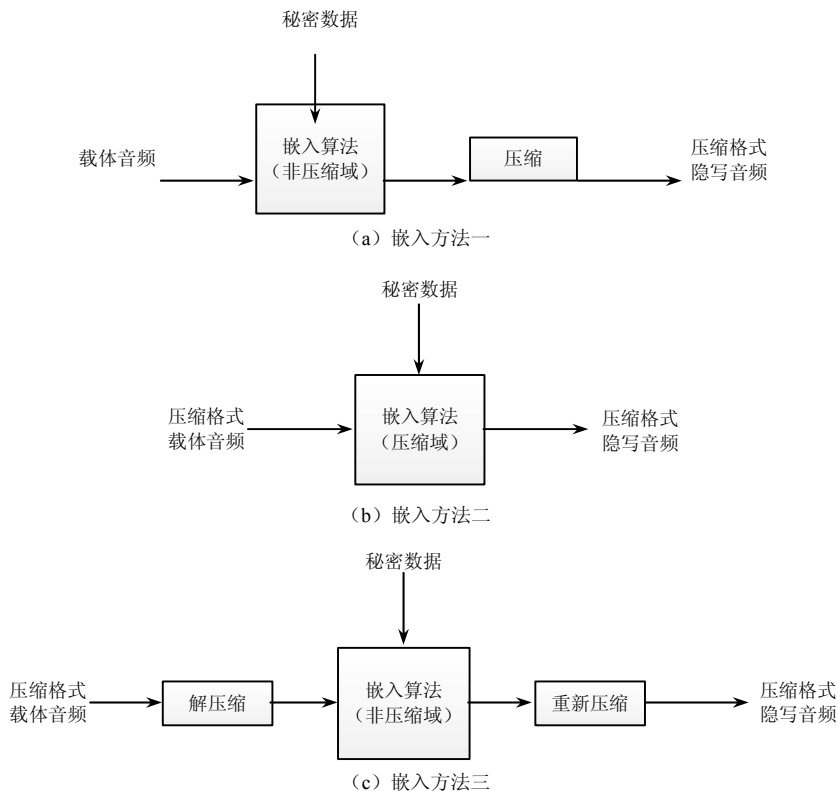


图 2.39 音频压缩域嵌入方法框图

目前在 MP3 中嵌入秘密信息最常用的是 **Mp3Stego** 软件，它是一款能够在 MP3 音频中隐藏秘密信息的隐写工具，嵌入与压缩过程同时进行。这种方法的鲁棒性较差，攻击者可以通过解压和重压缩来破坏秘密信息。另一方面，**Mp3Stego** 实际上并没有在音频压缩域中直接嵌入秘密数据的，而是在压缩过程中嵌入的，因此比较耗时。此外，还有学者提出一种直接在 MP3 压缩域中嵌入秘密数据的音频隐藏方法，该方法不是直接将秘密数据嵌入到频谱数据中，而是嵌入到比例因子上，这样不需要经过音频的压缩过程，比较节省时间，且克服了在 PCM 数据域上的算法在压缩过程中造成信息损失的缺点。还有的学者通过对 MP3 压缩过程中的 MDCT（修改 DCT 变换）部分系数进行调制来嵌入秘密信息，在压缩过程中同时嵌入秘密信息，嵌入效率较高，实验显示该方法能够抵抗编码率为 96kbps 以上的 MP3 压缩编解码攻击。还有的学者提出基于 Huffman 编码的 MP3 隐写算法，该算法直接在 MP3 编码中的 Huffman 码字上嵌入隐藏信息，通过改变部分 Huffman 码字达到嵌入秘密信息的目的。该算法不需要部分解码，具有不可感知、嵌入量大，计算量小的特点。

2.7 基于视频载体的隐写术

2.7.1 引言

前面两节讲了两种普遍应用的隐写载体，图像和音频。除此之外，隐写术最常用的载体是视频。限于篇幅，本节的内容只是给读者提供一个概况，具体方法留给读者去查

阅相关文献。对于视频中的隐写术来说,就是利用视频中存在的冗余数据来嵌入秘密消息。在这里,视频是载体,秘密消息同样可以是任意的比特流,隐写对象则是嵌入了消息的视频,视频中的隐写流程如图 2.40 所示。

发送方按照一定的消息嵌入算法和密钥将秘密消息(m)嵌入到载体视频(c)中,形成隐写视频(s)。为了满足信道带宽的需求,隐写视频在信道中传输时可能会被再次压缩,还可能遭受各种处理与攻击。接收方得到经过各种处理与攻击的隐写视频(s')后,按照消息提取算法和双方共享的密钥提取出秘密消息(m'),从而完成隐写通信用过程。

由于隐写术主要用于在保密通信时传输秘密消息,如军事通信以及情报或商业部门传递大容量的秘密文件。因此,视频中的隐写术主要具有如下特点。

(1) 不可感知性。包括视觉的不可感知以及统计的不可感知,这是对隐写算法的基本要求,即消息的嵌入对视频质量造成的影响对人的视觉系统来说是不可察觉的,同时消息嵌入后不能改变原始视频的统计特性,使得用统计方法无法检测出秘密消息。

(2) 安全性。由于视频所具有的大数据量以及帧间所具有的冗余,视频数据对于各种处理是非常敏感的,如帧添加、帧丢失、帧平均等。嵌入消息时必须考虑这些可能的处理,以提高安全性。

(3) 高容量。即在视频中能够嵌入更多的消息。与图像相比,视频由大量的帧序列组成,具有更大的载体空间,因此在其中嵌入消息具有更高的容量。

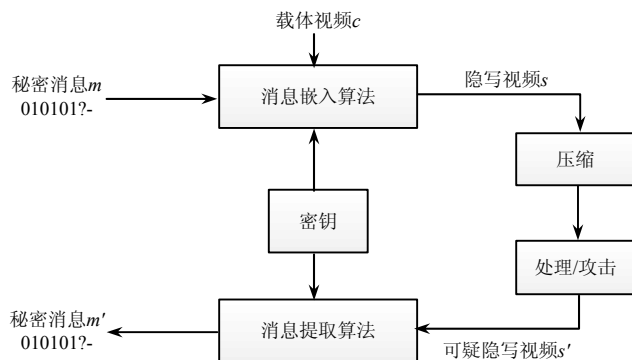


图 2.40 视频中的隐写流程

(4) 与视频压缩编码标准相结合。视频数据由于数据量大,在存储、传输时通常需要对其进行压缩。如果是在未压缩的视频中嵌入消息,由于是利用冗余数据来携带消息,而编码则需要去除冗余,若不考虑视频编码标准,则嵌入的消息很可能在编码过程中就会丢失。如果是在压缩视频中嵌入消息,很显然要与编码标准相结合。

视频水印技术的研究已成为热点,但视频隐写术研究还处于起步阶段,随着网络技术的发展和流媒体应用的不断增多,网络上的视频传输将越来越常见,视频中的隐写术将逐渐成为新的研究热点。目前,视频中的隐写术研究需要注意以下几个方面的问题。

(1) 由于视频主要以压缩的形式存在,因此视频中的隐写术应该重点考虑在压缩视频中嵌入秘密消息。虽然在压缩视频中嵌入消息可获得的嵌入容量较低,但是安全性较高,而且算法的复杂度更低。

(2) 进一步研究隐写算法的安全性问题以及安全性和容量之间的关系。由于数字视频的特点,在其中嵌入消息后,同一帧内部的不同区域之间以及不同帧之间的统计特性可能会有不同,从而会引起攻击者的注意。另外,视频传输时可能会经受的一系列视频

处理以及攻击等也是必须考虑的问题。

(3) 随着视频中的隐写术研究的不断深入,也要进行与之相对应的攻击技术的研究,即采用一定的技术手段检测、提取视频中嵌入的秘密消息或者使得秘密消息无效,亦即隐写分析技术。通过隐写分析,不仅可以发现隐蔽通信信道,阻止敌方的隐蔽通信,而且可以从中提取出秘密消息,获得有价值的情报。此外,隐写分析的研究还对安全的隐写算法的设计具有促进作用,从而可以提高隐写系统的安全性。它们之间既矛盾,又相互促进。

(4) 结合具体应用进行研究。如在视频会议系统中,通过将语音信号加密然后隐藏在图像信号中来进行传输,从而可以隐藏谈话的内容,增强通信系统的安全性。虽然视频中的隐写术研究没有图像中的研究那么广泛、深入,但是网络流媒体的迅速发展及应用需求的不断增多,视频中的隐写术将成为新的研究热点,相信在这个领域必将有更多更深入的研究成果不断出现。

2.7.2 未压缩视频中的隐写

这种情况是把消息直接嵌入到未经过压缩编码的原始视频中,然后再对嵌入了消息的视频进行编码。由于视频是由一系列的静止图像组成的,因此可以使用图像中的隐写方法。按照消息嵌入过程大致可以分为两类:一是空间域嵌入,即利用帧序列图像空间域的特性来嵌入消息,一般都是利用图像的冗余信息,最简单的情况是用待嵌入消息替换视频序列中的一些 LSB 位。只有知道嵌入位置才能提取消息,而且视频压缩编码要尽量去除冗余信息以压缩数据量,因此经过视频压缩后,嵌入的消息可能丢失,影响安全性。二是对帧序列图像进行一定的变换,在变换域进行消息的嵌入,这也是在未压缩视频中嵌入消息的常用方法。例如,我们可以将视频序列中的每一帧进行 8×8 块的 DCT 变换,然后通过对 DCT 系数的处理来嵌入秘密消息。2001 年,Abdulaziz 和 Pang 提出一种基于 DWT 变换和信道编码的消息嵌入方法^[29]。他们对待嵌入的图像秘密信息进行矢量量化,得到量化后的系数并进行信道编码。将视频帧序列变换到 DWT 域,将所选择的 DWT 系数用待嵌入系数代替,然后和未改变的 DWT 系数合并,形成新的系数,最后再进行反变换形成嵌入消息的视频帧。

与空间域方法相比,在变换域中嵌入消息可以把信号能量分布到空域的所有像素上,可以更好地与人的感知系统的某些特性相结合,有利于提高安全性,而且这些变换域的方法在图像隐写中的应用已经比较成熟。但是对于视频来说,如果不考虑相邻帧之间的相关性,而只是对各帧独立地进行处理,那么既没有充分利用视频信息的各种特性,同时嵌入的消息在传输过程中也容易通过帧的丢失以及添加等处理去除。我们可以通过增加冗余度和搜索对于帧处理的不变量来对抗这些攻击,把视频分割成时域的片断,每个片断由相似的连续帧组成,在段内每帧中嵌入相同的数据及帧同步索引,以对抗帧添加、帧丢失以及改变相邻帧次序等处理。该方法增强了安全性,但是会造成嵌入容量的浪费,而且增加了计算复杂度。

Pazarci 和 Dipcin 提出^[30]在视频编码之前,先进行置乱操作和消息嵌入,然后再对置乱并嵌入了数据的视频进行编码。该方法通过对块置乱参数进行预处理以产生块间差来嵌入数据,也可以增强嵌入消息的安全性。还有些方法结合现代信号处理的知识来进行,如 Joumaa 和 Davonie 提出^[31]一种基于独立分量分析(Independent Component Analysis, ICA)的隐写算法,他们对视频序列进行 ICA,从中提取出一系列统计独立源用于嵌入数据。

由于图像中的隐写算法比较成熟，可以充分利用这些算法在未压缩的原始视频序列中嵌入秘密消息，但是由于视频所固有的特点，需要对这些隐写算法进行重新设计。此外，由于原始视频中含有较多冗余，可以用于嵌入消息的空间更大，但是编码过程即可看作是对隐写算法的攻击，编码时可能会去除部分嵌入的消息。

2.7.3 压缩视频中的隐写

这种情况是在压缩码流中直接嵌入消息，或者把压缩码流解码后再嵌入消息。对于直接嵌入的情况，其显著特点是不需要对视频完全解码和再编码，但是由于是在压缩后的码流中嵌入消息，比特率的约束限制了消息的嵌入容量。因此，在压缩视频中直接嵌入消息在数字水印方面的应用较多。虽然隐写术不同于数字水印，但是其研究可以参考数字水印的相关方法。

在压缩视频中，也可以把秘密信息嵌入到 DCT 系数中。例如，我们可以利用扩频的思想在 MPEG-2 压缩视频的 DCT 系数中嵌入秘密信息。二维秘密信息经过扩展、放大和调制，得到一个随机序列，然后对其进行 8×8 的 DCT，并将 DCT 系数叠加到视频流的 8×8 的 DCT 系数上。在 DCT 系数中添加水印是常见的方法，算法成熟，而且可以借鉴图像中的方法。我们也可以修改压缩视频流的**可变长码**（Variable Length Coding, VLC）来嵌入秘密信息，而秘密信息的检测也是在 VLC 域中进行的，这样一来，检测速度较快，可以满足检测的实时性需求。由于 MPEG 视频序列中的大部分帧都是采用运动补偿预测技术来进行编码的，因此在运动矢量中也可以嵌入秘密信息。此外，MPEG-4 采用基于对象的编码方法的特点，因此我们也可以在视频对象（Video Object, VO）中嵌入秘密信息。这些算法基本上以数字水印技术的形式出现，将在第 3 章的视频水印技术这一节中介绍。除了 DCT 变换，还可以根据离散小波变换后的视频可以量化到位平面结构的特点，把位平面复杂度分割（BPCS）隐写方法应用到小波域来嵌入秘密消息。

对于将压缩视频解码后再嵌入消息的情况，按照解码的程度，又可分为完全解码和部分解码。例如，我们可以通过对压缩视频的部分解码来获得 DCT 系数，然后使用扩频方法进行秘密消息的嵌入，最后再重新编码完成嵌入过程。此外，我们也可以把视频分成一系列独立的由不固定数量的连续帧组成的语义场景，以场景为基本单元嵌入主要数据，同时把次要数据重复地嵌入具有相同场景的各帧内部，其实质是把部分数据嵌入压缩后的比特流中，部分数据嵌入在解码后的视频序列中。对于完全解码来说，也就是将视频解码成帧序列，然后按照未压缩视频中的方法嵌入秘密消息。

为了进一步提高视频传输的安全性以及秘密消息提取的准确性，还可以在消息嵌入之前，对待嵌入的秘密消息进行预处理，以提高其对抗攻击的能力，如对秘密消息进行置乱或错误校正编码等。

2.7.4 分析与比较

下面从隐写算法的性能、嵌入消息的安全性以及嵌入容量等方面对针对未压缩视频的以及压缩视频这两种情况下的隐写术进行分析与比较。

1. 隐写算法的性能

在数字水印应用中，水印的实时嵌入与快速检测是必需的，但是对隐写术来说，秘密消息的快速嵌入与检测虽不是最重要因素，但也必须考虑。在一些应用中，如在视频

会议系统中将谈话内容隐藏在视频中进行传输时，需要对话音信号进行快速的嵌入和提取，此时算法性能的优劣就显得尤为重要，而且由于视频的数据量较大，对其进行处理较为复杂，因此应该尽量减小隐写算法的复杂度，以确保秘密消息的快速嵌入与检测。

对于未压缩视频来说，可参照图像隐写方法嵌入消息，但是由于在嵌入消息后还需对其进行编码，因而算法的复杂度较高。对于直接处理压缩视频的情况，只需对比特流进行处理即可完成消息的嵌入，因而复杂度较低，而对于将压缩视频解码后再嵌入消息的情况，由于需要在嵌入消息后再对视频进行编码，因而相对来说，复杂度是最高的。

2. 鲁棒性和安全性

即秘密消息抵抗各种视频处理和恶意攻击的能力。这些视频处理包括帧丢失、帧添加、帧平均、改变相邻帧的次序、视频格式的转换或者改变视频码率等，是设计隐写算法时必须考虑的问题。

未压缩视频具有数据量较大和帧间冗余较多的特点，在传输之前通常要对其进行压缩，另外为了满足信道带宽的需求，在传输时可能会再次对其压缩，因此对于未压缩视频中的隐写，主要考虑如何使得嵌入的消息在后续的视频压缩后依然能够保存下来，如果嵌入算法设计不当，则嵌入的消息很可能在视频压缩时就会丢失。

对于压缩视频中的隐写术来说，由于是在压缩码流中嵌入消息，与视频编码标准的结合更紧密，因而抵抗视频处理的能力也更强，不过对于简单的诸如帧丢失、帧添加等视频处理还是需要考虑的。可以通过冗余嵌入，即在某个基于某种考虑划分出的视频段内重复嵌入消息来解决这一问题。

3. 嵌入容量

隐写术是利用载体数据中存在的冗余来嵌入秘密消息的，对于未压缩的视频序列来说，由于存在的冗余较多，因此获得的嵌入容量也较高，但是需要考虑在嵌入消息后还需再次对其压缩，这一过程可能会去除嵌入在其中的消息。

对于压缩视频来说，由于在压缩时已经去除了冗余，其中存在的冗余较少，一般都是结合具体的视频编码标准，利用 DCT 系数、可变长码、运动矢量以及视频对象等来进行秘密消息的嵌入，因而可获得的嵌入容量较低，但是对视频质量的影响也更小，一般用于视频水印。实际应用时，应根据具体的需求，选择合适的视频源作为载体，以在嵌入容量和安全性之间取得平衡。通过以上分析可以看出，在未压缩视频中嵌入消息可以获得较高的容量，但是需要考虑后续的视频压缩，嵌入消息的安全性相对较低；而对于压缩视频中的隐写来说，虽然嵌入容量较低，但安全性较高。在实际应用中，需要根据具体的应用需求选择合适的视频源，设计合适的隐写算法。

2.8 本章小结

本章首先概述保密通信的有关背景，接着概述隐写术的有关概念、分类和性能评价问题，然后按照载体类型的不同分别介绍基于文本、图像、音频和视频等载体的隐写术。由于文本的冗余空间比较低，而且涉及自然语言处理的知识，基于文本的信息隐写比以图像和视频为载体的信息隐写涉及更多的困难和挑战，因此基于文本的信息隐写相关成果相对较少。由于图像在 Internet 上的广泛使用以及图像中固有的大量冗余空间，使其成为一种最普遍的隐写术载体。虽然音频隐写面临着极大挑战性，但利用人类听觉系统存在的“掩蔽效应”，可将人类听觉系统无法感知的秘密信息嵌入到音频载体信息中。

在音频信息中隐藏机密信息，计算量小于在图像中的计算量。同时，随着多媒体技术的发展，音频信息在互联网上应用广泛，这些都使得音频隐写成为本领域的研究热点之一。数字视频由帧序列组成，具有信号空间大的特点，而且网络上的视频传输越来越常见，因此进行视频中的隐写术研究具有较大的理论与现实意义。

除了上述这四种载体外，还有许多其他类型的载体适用于隐写术：① 由于 FLASH 文件在互联网上的普及和广泛使用，使得通过 FLASH 文件为载体来传递秘密数据成为一个很好的选择。现在 FLASH 文件几乎存在于所有网站中，窃取者对含有秘密数据的 FLASH 文件不会产生太多怀疑，从而减少秘密数据被拦截或篡改的可能；② 网络通信中巨大的信息流通量使得用网络通信数据作为载体进行秘密通信的冗余量大，同时也使攻击者很难有足够的精力去检测网上所有的通信数据。为了保护网络中传递的秘密信息，以网络通信数据作为载体的保密通信是一个比较理想的方法；③ 相对于其他载体，由于网速的加快，网页在网络传输中不容易被截获和注意，更具有隐蔽性和安全性，如何利用网页安全传递重要信息和验证网页完整性和可信性成为网页信息安全重要内容；④ 此外，还可以在软件和文件系统中隐藏信息，例如可执行文件、NTFS 文件系统。



习题

1. 用 Matlab 或 C 语言编写一段程序，该程序能够打开一幅大小 256×256 的 256 灰度图像，抽取出其最低有效位平面，观察该位平面的特性（平均值和方差）。试着用给定的（可以随机生成）位平面（秘密信息）去替换该最低有效位平面，看看生成的隐写图像与原图像在视觉上有无差别？

2. 求下图 2×2 像素块 DCT 变换结果

30	44
53	26

题图 2.1 某 2×2 像素块

3. 用 Matlab 或 C 语言编写一段程序，该程序能够打开一幅大小 256×256 的 256 灰度图像，对该幅图像作 8×8 分块 DCT 变换，一共得到 64×64 块 DCT 系数块。为了在每块 DCT 系数块中隐藏 1 比特信息，用程序随机生成一段长度为 $64 \times 64 = 4096$ 的二值序列（每个元素要么是 0 要么是 1），利用调制阶距 $\Delta = 12$ 的抖动调制技术根据相应的秘密位修改每个 DCT 系数块的经过之字型排序后的第 8 个 AC 系数。对修改过的 DCT 系数块进行逆变换就可以得到隐写图像，观察隐写图像和原始图像之间的视觉差别，并计算该隐写图像和原始图像之间的峰值信噪比。

4. 求下图 4×4 像素块的二维 DFT 变换结果

3	44	11	18
26	26	13	13
0	42	13	36
34	10	4	33

题图 2.2 某 4×4 像素块

5. 设计一种基于同义词替换的文本隐写方法, 在文本“安倍上台后, 接连出访亚洲和欧洲国家, 但是至今没有访美, 这一方面可以认为安倍政权有意显示出其外交政策与小泉时代有所不同, 希望寻找日本外交更大的自主性; 另一方面, 日本与欧洲、澳大利亚的接近仍然还是在一定程度上配合美国。美国提出, 要构筑“远东地区的北约”, 应对包括中国崛起以及诸如朝鲜半岛局势等不确定因素。而这个所谓的“远东地区的北约”所涉及的国家也无非就是日本、澳大利亚等传统美国盟国。因此美国对于澳大利亚与日本的安全合作协议也大力予以推动。”中隐藏二值序列“1010010101101100”。

6. 用 Matlab 或 C 语言编写一段程序, 该程序能够打开一幅大小 256×256 的 256 灰度图像, 对该幅图像作整幅图像的 2D-FFT 变换, 实现一种修改其中 256 个中频系数以嵌入 256 比特信息的隐写算法。

7. 用 Matlab 或 C 语言编写一段程序, 该程序能够打开一幅大小 256×256 的 256 灰度 JPEG 图像, 利用 JSteg 隐写术选择每个 8×8 量化系数块的第 4 个 AC 系数进行秘密信息嵌入, 一共嵌入 128 比特信息, 比较顺序嵌入法和随机嵌入法得到的隐写图像以及原始载体图像的第 4 个量化 AC 系数的直方图。

8. 求题图 2.2 的二阶 Haar 小波变换结果。

9. 用 Matlab 测试一幅大小 256×256 的 256 灰度图像的三级 Haar 小波变换结果, 每次利用 LSB 替换隐写术在不同的中频子带嵌入相同的位平面, 观察对图像质量的影响情况。

10. 利用音频隐写术, 设计并用 Matlab 实现一种在一个音频文件中隐藏一幅灰度图像的隐写方法。载体对象为: 16bit、单声道、采样频率为 22kHz 的“Windows XP 启动.wav”音频文件, 大小为 166KB; 秘密信息: 8bit 的 128×128 的 Lena.bmp 文件, 大小为 17KB。

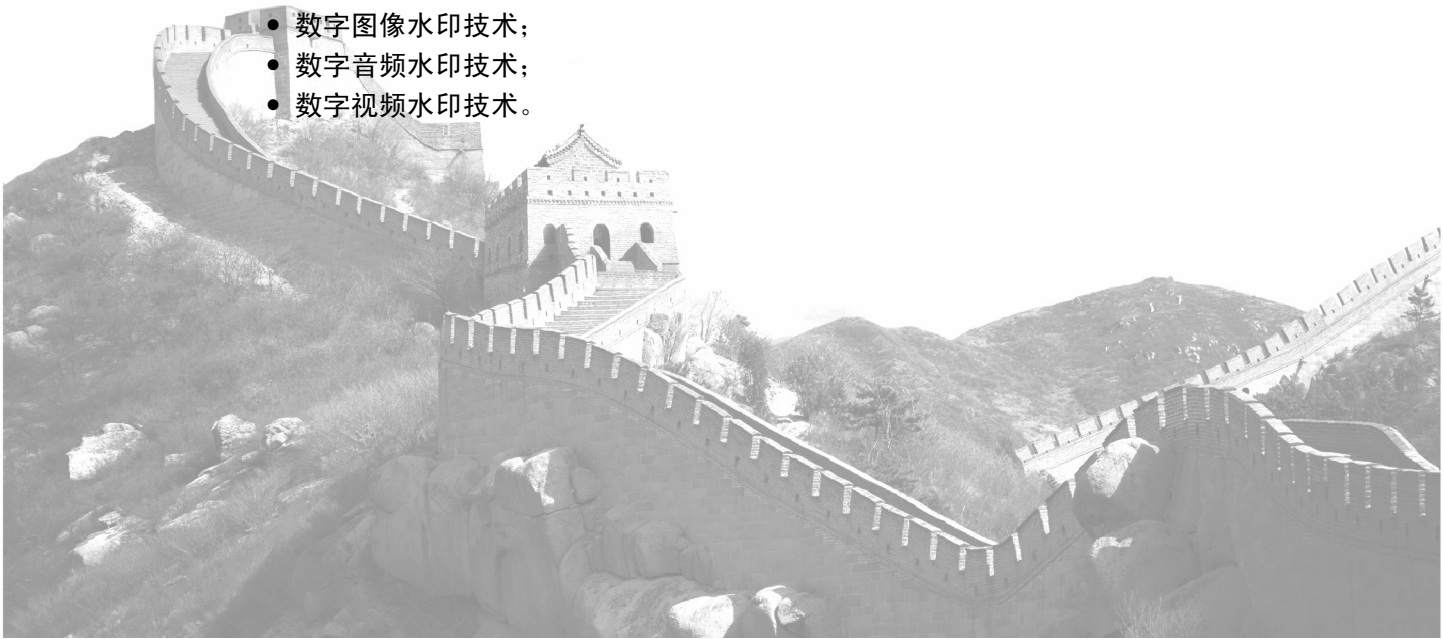
数字水印技术

本章引言

随着多媒体技术和网络技术的广泛应用，图像、音频、视频等多媒体内容的版权保护成为迫切需要解决的问题。数字水印技术（Digital Watermarking）作为版权保护的重要手段和一种新型的信息隐藏方法，近几年得到了迅速发展。数字水印技术是利用数字作品中普遍存在的冗余数据和随机性，把包含版权信息等内容的的数据（水印）嵌入到数字作品中，通过从加了水印的数字作品中检测或提取水印（有关版权的信息），从而起到保护数字作品版权的一种技术。数字水印技术区别于密码术的本质在于水印与载体数据是紧密结合并隐藏在其中的，水印是载体数据不可分离的组成部分。本章从数字水印技术的提出背景入手，首先介绍数字水印技术的相关概念、分类、框架模型和性能评价，然后按照载体类型的不同分别介绍用于图像、音频和视频等载体的数字水印技术。

本章重点

- 数字水印技术的相关概念和分类；
- 数字水印系统的框架模型和性能评价；
- 数字图像水印技术；
- 数字音频水印技术；
- 数字视频水印技术。



3.1 数字水印技术的提出背景

人们对数字水印技术研究兴趣的持续增长源于人们对版权保护问题的关注。近年来,随着计算机多媒体技术的迅猛发展,人们可以方便地利用数字设备制作、处理和存储图像、语音、文本和视频等信息媒体。与此同时,数字网络通信正在飞速发展,使得信息的发布和传输实现了“数字化”和“网络化”。在模拟时代,人们把磁带作为记录设备,盗版复制通常要比原始复制的质量低,而盗版复制的二次复制的质量更糟糕。而在数字时代,歌曲或电影的数字复制过程完全不损失作品质量。自从 1993 年 11 月因特网上出现了 Marc Andreessen 的 Mosaic 网页浏览器,因特网对用户变得友好起来,很快人们便开始乐于从因特网上下载图片、音乐和视频。对数字媒体而言,因特网成了最出色的分发系统,因为它不但便宜,而且不需要仓库存储,又能实时发送。如何在网络环境中实施有效的版权保护和信息安全手段,已经引起了国际学术界、企业界以及政府有关部门的广泛关注。其中,如何防止数字产品(如电子出版物、音频、视频、动画、图像产品等)被侵权、盗版和随意篡改,已经成为世界各国亟待解决的热门课题。

数字产品的实际发布机制的详细描述是相当复杂的,它包括原始制作者、编辑、多媒体集成者、重销者和国家官方等。图 3.1 给出一个简单的发布模型,“供应商”是版权所有、编辑和重销者的统称,他们试图通过网络发布数字产品 c 。“用户”也称为消费者(顾客),他们希望通过网络接收到数字产品 c 。“盗版者”是未授权的供应者,他们未经合法版权所有者的许可重新发送产品 c (如盗版者 A)或有意破坏原始产品(如盗版者 B)并重新发送其不可信的版本 c' 。从而消费者难免间接收到盗版的副本 c 或 c' 。盗版者对数字多媒体产品的非法操作行为,通常包括以下三种情况:① 非法访问:即未经版权所有者的允许从某个网站中非法复制或翻印数字产品;② 故意篡改:盗版者恶意地修改数字产品以抽取或插入特征并进行重新发送,从而使原始产品的版权信息丢失;③ 版权破坏:盗版者收到数字产品后未经版权所有者的允许将其转卖。

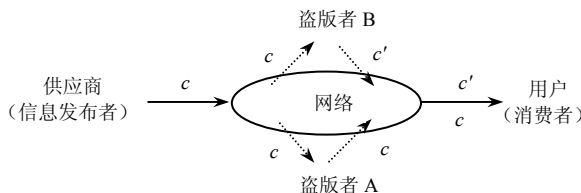


图 3.1 数字产品网络发布的基本模型

为了解决信息安全和版权保护问题,数字产品所有者首先想到的是加密和数字签名等技术。基于私用或公共密钥的加密技术可以用来控制数据访问,它将明文消息转换成旁人无法理解的密文消息。加密后的产品是可以访问的,但只有那些具有正确密钥的人才能解密。除此之外还可以通过设置密码,使得数据在传输时变得不可读,从而可以为处于从发送到接收过程中的数据提供有效的保护。数字签名是用“0”、“1”字符串来代替书写签名或印章,起到与书写签名或印章同样的法律作用。但这种数字签名在数字图像、视频或音频中的应用并不方便也不实际,因为在原始数据中需要加入大量的签名。另外,随着计算机软、硬件技术的迅速发展以及基于网络的具有并行计算能力的破解技术的日渐成熟,这些传统系统的安全性已经受到质疑。单靠通过增加密钥长度以增强保密

系统的可靠性已不再是唯一可行的办法。而且只有被授权持有密钥的人才可获得经过加密后的信息,这样就无法通过公共系统让更多的人获得他们所需要的信息。同时,一旦信息被非法破密,就没有任何直接证据来证明信息被非法复制和转发。因此,需要寻求一种不同于传统技术的更有效的手段,来保障数字信息的安全传输和保护数字产品的版权。

为了弥补密码技术的缺陷,人们开始寻求另一种技术来对加密技术进行补充,从而使解密后的内容仍能受到保护。数字水印技术有希望成为这样一种补充技术,因为它在数字产品中嵌入的信息不会被常规处理操作去除。数字水印技术一方面弥补了密码技术的缺陷,因为它可以为解密后的数据提供进一步的保护。另一方面,数字水印技术也弥补了数字签名技术的缺陷,因为它可以在原始数据中一次性嵌入大量的秘密信息。人们可设计某种水印,它在解密、再加密、压缩、数模转化以及文件格式变化等操作下保持完好。数字水印技术主要用于复制控制和版权保护。在复制控制应用中,水印技术可通过软件或硬件指出当前的复制行为是被禁止的。而在版权保护应用中,水印可用来标识版权所有,保证版税的合理支付。此外,水印技术还在一些其他场合得到应用,包括广播监控、交易跟踪、真伪鉴别、复制控制以及设备控制,这些将在最后一章中介绍。

数字水印技术是当前多媒体信息安全研究领域发展最快的热点技术,已经受到国际学术界和企业界的高度关注。数字水印技术是一门新兴的学科交叉的应用技术,它涉及不同学科领域的思想和理论,如信号处理、图像处理、信息论、编码理论、密码学、检测理论、概率论和随机理论、数字通信、对策论、计算机科学及网络技术、算法设计等技术,还包括公共策略和法律等问题。因此,无论从理论角度还是应用角度来看,开展对数字水印技术的研究,不但具有重要的学术意义,还有极为重要的经济意义。

3.2 数字水印技术的相关概念和分类

3.2.1 数字水印技术相关概念

提到水印,人们都会想到钞票中的水印。手持一张 20 美元的钞票,如果你观察带有 Andrew Jackson 总统肖像的一侧,在灯光下你将看见肖像中有一个水印显现。这个水印是在钞票制作过程中直接嵌入到纸币中的,因此人们很难伪造。它也阻止了假币制造者的一种常用伪造方法,即将 20 美元的墨洗掉后在同样的纸上打印上 100 美元。通常,钞票水印应该具有两条特性。首先,水印在通常情况下不可见,只有在特殊的观察条件下才变成可见(在这里,就是指将纸币放到光下)。其次,水印信息必须与载体对象相关(在这里,水印用来表示纸币的真实性)。

除了钞票,水印可以用于其他物理对象甚至电信号中。织物,衣服商标以及产品包装都是一些具体的可以用特制的染料和墨水加入水印的实例。音乐、照片和视频所代表的电子媒体则是一些常见的可嵌入水印的信号类型。本书只关注电子信号的水印技术,并且使用下列术语来描述这类信号。

作品(或产品):具体的一首歌、一段视频、一幅图画或者它们的一个复制。不含水印的原始作品常被称作“载体作品”。嵌入水印后的作品称为“含水印作品”。

内容:所有可能的作品集合。比如,音乐是一类“内容”,而具体的某首歌则是一件作品。

媒体:用来再现、传输和记录“内容”的媒介。

数字水印技术(Digital Watermarking)是一种信息隐藏技术,它的基本思想是在数

字图像、音频和视频等数字产品中嵌入秘密信息以便保护数字产品的版权、证明产品的真实可靠性、跟踪盗版行为或者提供产品的附加信息。其中的秘密信息可以是版权标志、用户序列号或者是产品相关信息。一般，它需要经过适当变换再嵌入到数字产品中，通常称变换后的秘密信息为**数字水印**（Digital Watermark），在诸多文献中论及了各种形式的水印信号。通常，可以定义水印为如下的信号 w

$$w = \{w_i \mid w_i \in \Omega, i = 0, 1, 2, \dots, L-1\} \quad (3.1)$$

式中， L 为水印序列的长度， Ω 代表值域。实际上，水印不仅可以为一维序列，也可以是二维阵列，甚至是三维和高维信号，这通常要根据载体对象的维数来确定，如音频对应一维，静止图像对应二维，动态图像对应三维。本书为了描述方便，通常用式 (3.1) 表示水印信号，对于高维情况，相当于将高维信号按一定顺序展成一维形式。水印信号的值域可以是二值形式，如 $\Omega = \{0, 1\}$ 、 $\Omega = \{-1, 1\}$ 或 $\Omega = \{-r, r\}$ ，或者是高斯白噪声（如均值为 0，方差为 1 的高斯白噪声 $N(0, 1)$ ）等其他形式。

3.2.2 数字水印技术和隐写术的区别

上一章所述的隐写术是一种保密通信技术，数字水印则主要用于数字产品的版权保护及其真实性和完整性认证。隐写术和数字水印技术的基本思想都是将秘密信息隐藏在载体对象中。但是，隐写术和数字水印之间有差异。对于隐写术来说，所要发送的秘密信息是主体，是重点保护对象，至于用什么载体对象进行传输无关紧要。对于数字水印技术来说，载体对象通常是数字产品，是版权保护对象，而所嵌入的信息则是与该产品相关的版权标志、购买者或者其他相关信息。例如，奴隶头上文着消息“此奴隶为 Histiaeus 所有”，即这条消息提到了这个奴隶（载体作品）的所属关系。如果别人企图宣称拥有这个奴隶的所有权，Histiaeus 便可剃去此奴隶的头发以证明自己是他的主人。在此情况下，此奴隶（载体作品）对 Histiaeus 而言是最有价值的，而那则消息只是提供了此作品的有用信息。由此可见，在媒体作品中嵌入消息的系统可分两类：一类是数字水印系统，其中嵌入的消息与载体作品有关；另一类则是非水印系统，其中嵌入的消息同载体作品无关。与此同时，也可将其分为隐写系统和非隐写系统两类。在前者中，消息的存在性是保密的；而在后者中，消息的存在性并不需要保密。因此，可将信息隐藏系统分为四类：隐写式数字水印系统、非隐写式数字水印系统、非水印的隐写系统和非水印非隐写系统，下面通过例子说明每一类的含义。

1. 隐写式数字水印系统

在 1981 年，英国内阁秘密文件的影印版居然出现在报纸上。为了确定泄密者，撒切尔决定给每个内阁成员散发可被独立辨认的文件复制。每个复制都具有不同的单词间距，用以对接收者的身份进行编码。这样，泄密者便很容易被辨认出来。这就是**隐写式数字水印**的一个例子，其中隐藏的格式便是用来对每份复制接收者进行编码的水印；同时它也是隐写式的，因为部长们并不知道水印信息的存在。

2. 非隐写式数字水印系统

俄国圣彼得堡的 Hermitage 博物馆将其大量名贵收藏的高质量数字复制搬上了主页。其中每幅画中都嵌入了 Hermitage 博物馆版权所有的水印信息，并在每个网页的底部附有作品已经嵌入水印的说明，以及图像不可复制的警告。这种宣称每幅图像均已嵌入水印的说明就是**非隐写式数字水印**的一个例子，有助于阻止盗版。

3. 非水印的隐写系统

在载体作品中秘密地嵌入同载体作品无关的数据是一种**非水印的隐写**过程，它一直为军方所关注。例如，Simmons 描述了一个引人注目的隐蔽信道的例子，涉及检验美国和苏联之间 SALT-II 条约执行情况的有关问题。SALT-II 条约允许两国拥有较多的导弹发射井，但要对导弹数量进行限制。根据条约，两国均要在导弹发射井中安装由第三国提供的传感器。每个传感器都会向第三国报告发射井是否被占用的信息，除此之外不传递其他任何信息。但是，第三国可能会将传感器设计成能够在合法消息的掩盖下传输其他附加信息，如发射井的位置等。

4. 非水印非隐写系统

非水印非隐写系统公开地在载体信号中嵌入同载体信号无关的辅助性隐藏信息。在 20 世纪 40 年代末期，在某一频段（如 800Hz）的广播中嵌入一个时间编码是普遍做法。这种时间码呈周期性嵌入，如每隔 15 分钟。这种编码在广播中被人的听觉所掩盖，但它并不是水印，因为这个消息（当前时间）同广播内容无关。另一方面，它也不是隐写式的，因为只有在其存在性已知情况下这种时间码才有用。

依据嵌入数据是否同载体作品相关，可以找出各类数据隐藏方法的不同应用。实际上，水印技术同那些非水印系统相比存在很多相似的地方，甚至在某些方面完全一样。因此，数字水印技术和隐写术之间许多技术是互相适用的。

3.2.3 数字水印及数字水印技术的分类

数字水印是加在数字图像、音频或视频等媒体中的信号，这个信号使人们能够建立产品所有权，辨识购买者或提供数字产品的一些额外信息。从含水印图像中的水印是否可见分为可见水印和不可见水印两大类。本书主要讨论不可见水印，故下面的论述中如没有特殊申明，都是指不可见水印。从水印生成是否依赖于原始载体来分，可分为非自适应水印（独立于原始载体的水印）和自适应水印。独立于原始载体的水印可以是随机产生的、用算法生成的，也可以是事先给定的，而自适应水印是考虑原始载体的特性而生成的水印。从含水印载体的抗攻击能力即鲁棒性来分，可分为脆弱水印、半脆弱水印和鲁棒水印。脆弱水印对任何变换或处理都非常敏感，半脆弱水印是对一部分特定的图像处理有鲁棒性而对其他处理不具备鲁棒性。鲁棒水印对常见的各种图像处理方法都具备鲁棒性。从水印检测是否需要原始图像参与来分，可分为明检测水印（私有水印）和盲检测水印（公有水印）。私有水印的检测需要原始图像的参与，而公有水印不需要原始图像的参与。根据水印的应用目的不同，可分为：版权保护水印、篡改提示水印（内容认证水印）、版权跟踪水印（数字指纹）、复制控制水印、标注水印（用来注释载体的拍摄日期等）和隐藏通信（保密通信）水印等。

相应地，数字水印技术也可以分为两大类：可见水印技术和不可见水印技术。本章讨论不可见水印技术。不可见水印技术主要可以分为时/空间域、变换域和压缩域三种。时空域水印技术是用各种各样的方法直接修改载体的时/空域采样（如直接修改像素的最低位）。这类算法的鲁棒性不高，且能够嵌入的水印信息不太多，否则从视觉上能看出来。而变换域水印技术是对原始载体进行各种各样的变换后嵌入水印，如 DCT、DFT、DWT 等。压缩域水印技术是指在 JPEG 域、MPEG 域等压缩域内进行的水印嵌入，这类算法直接对相应的压缩攻击具有鲁棒性。有些学者将公钥密码体制借鉴到水印系统中，使检测密钥与嵌入密钥不同，这种水印系统称为公钥水印系统，否则称为私钥水印系统。根

据水印提取后原始载体能否无失真恢复可将水印系统分为可逆水印系统和不可逆水印系统两大类。根据原始载体的不同,可将水印技术分为:音频水印技术、图像水印技术、视频水印技术、三维目标或三维图像水印技术、文档水印技术、数据库水印技术、集成电路水印技术和软件水印技术(加在程序代码或可执行文件中的水印)等。根据水印技术是否利用自适应技术(包括在生成和嵌入过程中,包括嵌入参数和嵌入位置的自适应),可将数字水印系统分为自适应数字水印系统和非自适应数字水印系统两大类。此外,有些学者还提出非线性数字水印技术(基于混沌、分形、神经网络、遗传算法)、第二代数字水印技术(基于不变特征点)、多功能水印技术(同时嵌入多种功能的水印)等概念。

3.3 数字水印系统的框架模型

3.3.1 数字水印系统基本框架

粗略来看,数字水印系统包含嵌入器和检测器两大部分。嵌入器至少具有两个输入量:一个是原始信息,它通过适当变换后作为待嵌入的水印信号;另一个就是要在其中嵌入水印的载体作品。水印嵌入器的输出结果为含水印的载体作品,通常用于传输和转录。之后,这件作品或另一件未经过这个嵌入器的作品可作为水印检测器的输入量。大多数检测器试图尽可能地判断出水印存在与否,若存在,则输出为所嵌入的水印信号。图 3.2 给出了数字水印系统基本框架的详细示意图。它可以定义为九元体 $(M, C, W, K, Ge, Em, At, De, Ex)$, 分别定义如下。

(1) M 代表所有可能原始信息 m 的集合;

(2) C 代表所要保护的数字产品 c (或称为作品) 的集合, 即内容;

(3) W 代表所有可能水印信号 w 的集合;

(4) K 代表水印密钥 k 的集合;

(5) Ge 表示利用原始信息 m 、密钥 k 和原始数字产品 c 共同生成水印的算法, 即

$$Ge: M \times C \times K \rightarrow W, \quad w = Ge(m, c, k) \quad (3.2)$$

需要说明的是原始数字产品不一定参与水印生成过程, 因此图 3.2 中用虚线表示。

(6) Em 表示将水印 w 嵌入数字产品 c 中的嵌入算法, 即

$$Em: C \times W \rightarrow C, \quad s = Em(c, w) \quad (3.3)$$

这里, $c \in C$ 代表原始产品, $s \in C$ 代表含水印产品。为了提高安全性, 有时在嵌入算法中包含嵌入密钥。

(7) At 表示对含水印产品 s 的攻击算法, 即

$$At: C \times K \rightarrow C, \quad c' = At(s', k') \quad (3.4)$$

这里, $k' \in K$ 表示攻击者伪造的密钥, s' 表示被攻击后的含水印产品。

(8) De 表示水印检测算法, 即

$$De: C \times K \rightarrow \{0, 1\}, \quad De(s', k) = \begin{cases} 1 & \text{如果 } s' \text{ 中存在 } w & (H_1) \\ 0 & \text{若 } s' \text{ 中不存在 } w & (H_0) \end{cases} \quad (3.5)$$

这里, s' 表示可疑的含水印产品, H_1 和 H_0 代表二值假设, 分别表示水印的有无。

(9) Ex 表示水印提取算法, 即

$$Ex: C \times K \rightarrow W, \quad w' = Ex(s', k) \quad (3.6)$$

这里, w' 表示提取出来的水印。

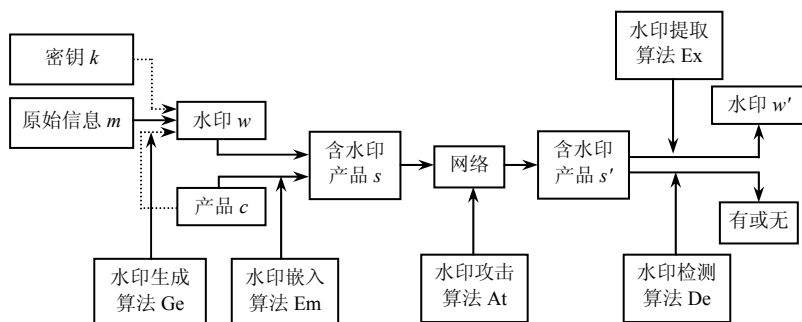


图 3.2 数字水印系统基本框架

3.3.2 基于通信系统的数字水印模型

从本质上说，数字水印技术是一种通信，即在水印的嵌入者和接收者之间传递一条消息。很自然，人们试图用传统的基本通信系统模型来表示整个水印系统。Cox 首先提出了三种通信模型^[5,8]。这三种模型之间的差异在于如何将载体作品融入到传统的通信模型中。在第一种基本模型中，将载体作品完全看作噪声。在第二种模型中，载体作品仍然被看作噪声，但该噪声作为**附加信息**输入到信道编码器中。在第三个模型中，载体作品不看作噪声，而看作第二个信息，这个信息和原始消息一起以多路复用形式进行传输。

1. 基本模型

图 3.3 和图 3.4 给出两种基本的数字水印系统通信模型，其中图 3.3 采用了**明检测器**，而图 3.4 采用了**盲检测器**。在这两种模型中，水印嵌入器被看成信道，输入消息通过信道进行通信，而载体作品是信道的一部分。为描述方便，这里称水印生成算法为水印编码器，并将水印编码器并入到水印嵌入器中。不管采用明检测器还是盲检测器，嵌入过程的第一步都是将消息 m 映射成一个和原始产品 c 具有相同类型和维数的嵌入模式 w_a ，这实际上是一个水印生成过程。例如，在图像中嵌入水印（空域），水印编码器（水印生成器）将产生一个和目标对象大小相同的二维图像模式。而在音频信号中嵌入水印（时域）时，水印编码器将产生一个相同长度的一维音频模式。这种映射通常需要借助水印密钥 k 来完成。嵌入模式是通过几个步骤计算出来的：① 预先定义一个或多个参考模式（用 w_r 表示，比如说伪随机序列或混沌序列），它们依赖于某一个密钥 k ；② 这些参考模式结合起来产生一个对消息 m 进行编码的模式，通常称之为消息模式 w ，本书称之为待嵌入水印 w ，是水印生成算法的输出；③ 将这一消息模式按比例缩放或修改以产生嵌入模式 w_a 。图中的水印编码器都没有考虑载体作品这个因素，人们把这种类型的生成器称为非自适应生成器。嵌入模式 w_a 嵌入到作品 c 中后得到含水印作品 s ，该作品将经受某种方式的处理，其处理的效果等价于噪声 n 的叠加。这里，对作品的处理有可能包括压缩和解压缩，模数转换，信号增强等无意攻击，也可能包括试图去除水印等恶意攻击行为。注意到所有这些处理都依赖于含水印作品，故为简便起见用加性噪声来等价这些处理。

图 3.4 中的水印检测器和水印解码器并没有实质性的区别。若使用图 3.3 的明检测器，那么检测过程将由两步组成：首先，从接收到的作品 s' 中减去原始载体作品 c ，从而获得含噪声的水印模式 w' 。接着，由水印解码器利用水印密钥来进行解码。由于载体作品在嵌入器中的叠加已经由检测器中的减法所抵消，故 w_a 和 w' 之间的差异实质上是由噪声引起的，从而可忽略载体作品的影响。这就意味着水印编码器、噪声叠加和水印解码

器一起组成了一个与基本通信模型类似的系统。在一些更先进的明检测系统中，检测时不需要全部原始载体作品，而利用 c 的某个函数，通常是一个数据简化函数，来抵消嵌入端由于载体作品叠加所产生的“噪声”影响。例如，某些检测器只需利用原始图像 DCT 变换系数中的一小部分。在图 3.4 的盲检测器中，由于不需要原始载体作品参与检测，故不需要在解码之前减去原始载体。在这种情况下，原始载体作品和攻击的组合可看成单个噪声。接收到的含水印作品 s' 可看成嵌入模式 w_a 遭到破坏后的一个作品版本，且整个水印检测器可看成信道解码器。

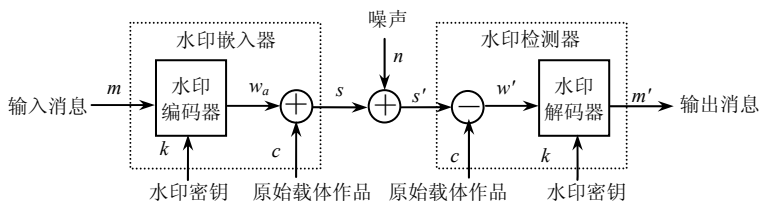


图 3.3 明检测数字水印系统映射成通信模型

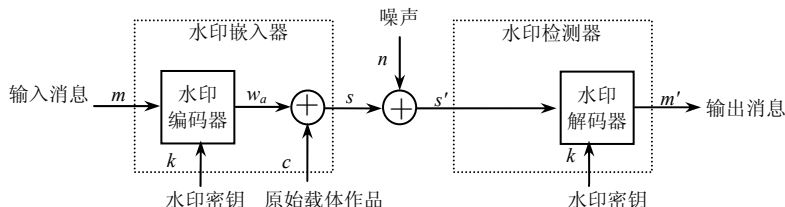


图 3.4 盲检测数字水印系统映射成通信模型

在交易跟踪或版权保护应用中，人们希望检测到的信息与已嵌入信息相同的概率最大，这与传统通信系统的目标是一致的。然而需要注意的是，在进行认证的应用场合中，由于其目的不是传递一条消息，而是检验作品在水印嵌入之后是否被修改或如何被修改，故通常不采用图 3.3 和图 3.4 的模型来表征认证系统。

2. 基于带有附加信息的通信系统的水印模型

图 3.3 和图 3.4 这两个模型并不适用于所有可能的数字水印系统，因为它们都认为待嵌入水印与载体作品无关。然而，对于嵌入器来说，原始载体作品 c 显然是已知的，故没有理由在数字水印生成过程中禁止使用原始载体。实际上，在水印生成过程中考虑原始载体的特性往往可提高含水印作品的保真度。如果在嵌入模式 w_a 生成中允许水印编码器利用 c ，那么人们称该生成器为自适应生成器。为此，图 3.5 给出了另一种水印模型，它允许 w_a 依赖于 c 。这个模型与图 3.4 几乎完全相同，唯一区别在于 c 被看成水印编码器的一个附加输入。应该注意，这一变化将允许嵌入器将 s 置为任意理想值而只需简单地让 w_a 满足 $w_a = s - c$ 。如果继续将载体作品看成信道中的噪声 ($c + n$) 的一部分，那么这个模型就是一个能在发送端提供附加信息的通信系统的例子，换句话说，就是嵌入器能够利用有关信道噪声的一些信息，特别是 c 本身。这带有附加信息的通信系统最早由香农研究，最近一些学者已经开始将带有附加信息的通信系统概念应用于水印。需要指出，图 3.5 中所示的为盲检测器，也可以采用图 3.3 所示的明检测器，只需将原始载体作品作为检测器的一个附加输入。

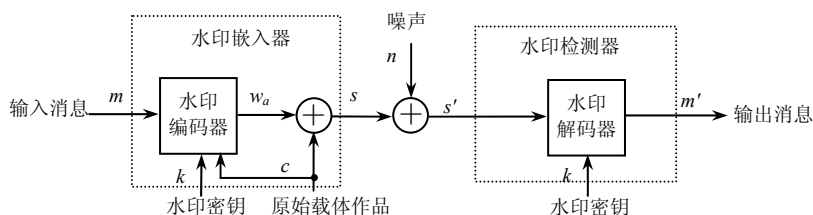


图 3.5 基于带有附加信息的通信系统的水印模型

3. 基于复用通信系统的水印模型

图 3.6 给出一种基于复用通信系统的水印模型。这里，不再将载体作品看成信道中的一部分，而是看成存在于同一信号 s 中并与水印信息一起传送的第二条消息。这两条消息将分别由两个完全不同的接收器来检测和解码：一个是人，另一个是水印检测器。水印嵌入器将 m 和 c 合并成单一信号 s 。这种合并与传统通信系统中多个信息通过时分复用、频分复用或是码分复用的方式在一个信道进行传输的情况类似。然而两者存在这样一个区别：在传统通信系统中，对于不同消息来说所采用的基本技术是相同的，而且通过一个单一的参数（时间、频率或编码序列）将多条消息分开。相比之下，在水印系统中是通过不同的技术将两条消息分开：水印检测与人类感知。这一点就如同对于一条消息来说采用频分技术，而对于另一条消息则采用扩频编码。

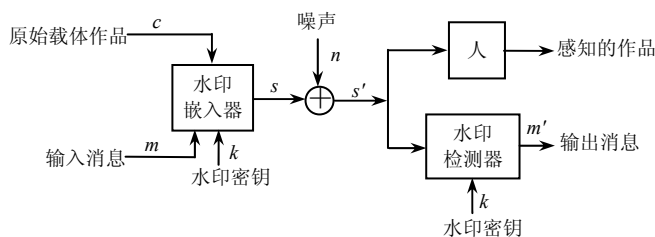


图 3.6 基于复用通信系统的水印模型

当信号通过信道之后，它一方面进入人类感知系统，另一方面进入水印检测器。当对 s' 进行观察时，观察者应该在不受水印影响的条件下感知与原始载体作品相近的作品。当对 s' 中的水印进行检测时，检测器应该包含原始的水印信息，并不受载体作品的影响。如果水印检测器是明检测器，那么它将把原始的载体作品或有关载体作品的某一个函数作为第二个输入。该框图强调了水印与载体作品之间的对称性。这里，需要强调术语“信噪比”的两种不同用法。当讨论作品逼真度时，“信号”指的是载体作品，而“噪声”指的是水印。当讨论水印系统的有效性和鲁棒性时，则“信号”指的是水印，而“噪声”则指的是载体作品。一般来说，从上下文中可以很清楚地看出这个术语想要表达什么意思。在水印嵌入算法中，通过在设计嵌入模式之前对载体模式进行检验，能够解决盲水印嵌入的有效性问题。类似地，在考虑逼真度问题时，人们可以先用一个感知模型来检验水印如何影响作品，然后通过调整 w_a 来尽可能减少这种影响，进而改进整个系统。

3.3.3 数字水印系统的几何模型

上一小节所描述的各种模型使人们能够在设计与分析水印系统时参考通信领域的知识。这一小节从另一个角度，即利用几何学的观点来认识水印系统。为了从几何学角度认识水印，首先需要想象出一个高维空间，该空间中的每一个点都对应于一个作品。人

们将该空间称为**媒体空间** (Media Space)。由于媒体空间维数高而复杂, 为了简化, 人们通常希望考虑媒体空间的投影或变形, 并将这样的空间称为**标记空间** (Marking Space), 即水印的嵌入空间。然后, 可以根据媒体或标记空间的不同区域和概率分布来认识水印系统, 主要包括下面一些内容。

(1) **未嵌水印作品的分布**表明每一件原始作品的出现概率。

(2) **可接受的逼真度区域**表示在这一区域中所有的作品从本质上来说与给定的载体作品是一致的。

(3) **检测区域**用来描述检测算法的性能。

(4) **嵌入分布或区域**用来描述嵌入算法的效果。

(5) **失真分布**表明在正常使用中作品有多大可能会被扭曲变形。

作品可以看成是一个 N 维媒体空间中的点。而媒体空间的维数 N 就是用来表示每一件作品的采样点数。以单色图像为例, 这个维数就是像素个数; 对于用三种色彩分量 (红、绿、蓝) 表示的图像, 维数 N 就是像素个数的 3 倍; 对于音频信号, N 就是采样点数; 而对于视频信号, N 就是帧数乘以每一帧中的像素数 (如果视频信号是真彩色的, 那么还需再乘以 3)。通常只对数字媒体感兴趣, 故每一个采样都被量化并限制在一定的范围内。例如, 8 位灰度图像中的每个像素值都被量化成 0~255 之间的一个整数。这意味着存在一个所有可能作品的有限 (虽然很大) 集合, 并且这些作品被排列在媒体空间的一个长方形点阵中。格点之间或边界之外的点并不对应于那些能够以数字形式表示的作品。然而, 通常情况下量化步长足够小而边界足够大, 以至于人们常常掩盖上述事实而假设媒体空间是连续的 (也就是说, 空间中的所有点, 甚至是格状结构之外的点, 也都对应于可以用数字形式实现的作品)。下面的各部分分别讨论媒体空间中每一种不同的概率分布和区域, 最后介绍标记空间的有关概念。

1. 不含水印作品的分布

不同的作品进入水印嵌入器或检测器的概率是不同的。在音频信号中, 人们更有可能将水印嵌入音乐中而不是纯粹的静音中。在视频信号中, 人们更有可能将水印嵌入自然景物的图像中而不是在视频作品的“雪花”中。由于音乐和自然图像具有不同的统计分布, 故在水印系统中必须考虑这些不同的分布特性。当对水印系统的特性 (如虚检概率和有效性等) 进行评估时, 很重要的一点就是对水印系统所要处理内容的先验分布进行建模。这可以通过所有格点的概率分布或媒体空间中所有点的概率密度函数来表示。**不含水印的作品**的分布情况存在很多统计模型。最简单的模型就是椭圆高斯分布。通过使用拉普拉斯分布或一般的高斯分布可以获得大多数媒体的更为精确的模型。最复杂的模型则是尝试将作品描述成随机变量。应该注意到, 未嵌水印作品的分布是依赖于应用的。例如, 卫星图像就是从一个与新闻图片完全不同的分布中获得的。音乐也是从一个与语音完全不同的分布中获得的。这种对于不同类型作品具有不同分布的多样性将会产生一些问题。如果人们使用一种分布来估计水印检测器的虚检概率, 然后将这一检测器用于另一个应用场合, 那么估计将不准确。在一些需要具有非常低的虚检概率的应用中 (如版权控制), 这个问题将会变得非常严重。

2. 具有可接受逼真度的区域

设想仅对一幅原始图像的单个像素改动一个亮度单位。这样一来, 将产生一幅新图像 (媒体空间中的一个新矢量), 然而从感知程度上来说它与原始图像是没有区别的。很显然, 在媒体空间中存在许多这样的图像, 并且可以想象出原始载体 c 附近的一个区

域, 在这个区域中每一个矢量都对应一幅在感知上与 c 没有区别的图像。通常将媒体空间中这些与载体作品 c 在实质上并无差别的矢量所构成的区域称为**可接受逼真度的区域**。对于一个给定作品 c , 确定其可接受逼真度的真正区域是很困难的, 因为人们对人类感知所了解的知识太少。通常情况下人们是通过感知距离设置一个门限值来近似估计这一区域。例如, 通常用均方误差 (MSE) 作为一个简单感知距离度量, 定义为

$$d(c_1, c_2) = \frac{1}{N} \sum_{i=1}^N (c_{1i} - c_{2i})^2 \quad (3.7)$$

其中 c_1 和 c_2 分别是媒体空间中的 N 维矢量。如果对这个函数设定一个界限 T , 那么可接受逼真度的区域就是一个 N 维超球体, 半径为 $(NT)^{0.5}$ 。更为复杂的距离函数能够对于人的判决给出更好的预测, 这些函数常常以 **JND** (Just Noticeable Difference) 为单位来度量感知距离, 有关感知距离函数的内容可参阅 Watson 于 1993 年发表的文献^[32]。

3. 检测区域

对于给定消息 m 和水印密钥 k 来说, **检测区域**就是媒体空间中的一个子集, 该子集中的所有作品可由检测器检测出消息 m 的存在。像可接受逼真度的区域一样, 检测区域常常 (但并不总是) 由一个门限值来定义。用于度量检测器输入和参考模式 w_r 之间的相似度称为检测测度。如最常见的检测测度就是线性相关 $e(s, w_r)$ 。为了找到对应这个检测器的检测区域形状, 首先要注意含水印作品 s 和参考模式 w_r 之间的线性相关公式如下

$$e = \frac{1}{N} \sum_{i=1}^N s_i \cdot w_{ri} \quad (3.8)$$

它等于它们的欧几里得长度乘积乘以它们之间夹角的余弦再除以 N 。因为 w_r 的欧几里得长度是不变的, 故该测度就等于寻找矢量 s 在矢量 w_r 上的正交投影。对于满足线性相关值大于 T 的所有点的集合就是与 w_r 垂直的平面一侧所有点的集合。要使一条水印信息能成功嵌入到载体中, 必须使含水印作品 s 位于具有可接受逼真度的区域和检测区域之间的重叠部分, 如图 3.7 所示。该图表示了一个线性相关盲检测水印系统的二维模型, 其中可接受逼真度的区域建立在均方误差基础上, 而检测区域则基于线性相关。

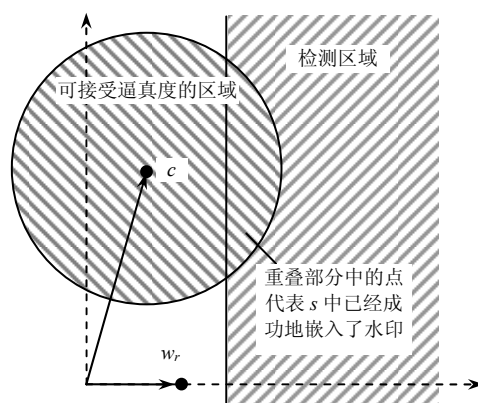


图 3.7 一个线性相关盲检测水印系统的可接受逼真度的区域和检测区域

4. 嵌入分布或区域

一个水印嵌入器就是一个函数, 它将一件作品、一条消息还有可能加上一个密钥映射成一件新的作品。一般来说这个函数是一个确定性函数, 这样对于给定的原始作品

c , 消息 m 和密钥 k , 嵌入器总是输出相同的含水印作品 s 。然而, 原始作品是从未嵌入水印作品的分布中随机选取的, 从而嵌入器的输出可看成是随机的。由此可见, 从嵌入器的输出给定作品 s 的概率就是从未嵌入水印作品的分布中获取 c 的概率。如果是由几个未嵌入水印的作品映射成 s , 那么获得 s 的概率就是获得这几件作品的概率之和。人们将这一概率分布称为**嵌入分布**。

5. 失真分布

为了判断对于水印内容的攻击效果, 人们需要知道如下这个概率, 即在给定未受攻击的含水印作品 s 情况下得到某一指定的受攻击作品 s' 的概率。这一条件概率的分布与传统通信理论中描述传输信道特性所使用的分布属同一类型, 这里将这一分布称为 s 附近的**失真分布**。

在水印算法分析中, 经常假设失真分布可被模型化为加性高斯噪声, 从而大大简化分析过程, 但它并不完全符合现实中的模型。对于数字内容的处理很少会产生高斯噪声效果。实际上, 它们中只有非常少的一部分是随机的。在正常的使用过程中, 作品内容在更大概率上会经历有损压缩、滤波、去噪、空间或几何变形这样的失真。一般来说, 总存在一些确定性函数可用来对作品内容进行处理。因此, 由于作品失真所产生的“噪声”在很大程度上依赖于作品内容本身。

6. 标记空间

对于一些简单的水印系统, 例如线性相关盲检测系统, 在媒体空间中确定嵌入和检测区域并不是很困难。然而, 大多数的应用需要更为复杂的算法, 这些算法在媒体空间中分析起来非常困难。当考虑这种系统时, 通常可将系统的一部分看成将媒体空间映射或变换到**标记空间**。系统的其余部分则被看成一个较为简单的水印系统, 它是对标记空间而不是媒体空间操作。

设计水印检测器时常常带有明确的有关标记空间的想法。如图 3.8 所示, 该检测过程包括两个步骤: 第一步是**特征提取**, 它对作品内容进行预处理, 例如频域变换、滤波、块平均、几何或时间定位和特征提取等。处理结果是得到一个矢量, 它是标记空间中的一个点, 并且其维数有可能比原来的要小, 将这个矢量称为**提取标记**。第二步是确定提取标记中是否包含水印, 如果有, 则对嵌入信息进行解码。这通常需要将提取标记与一个或多个预先定义的标记进行比较。第二步可想象成一个简单的水印检测器对标记空间中的矢量进行处理。

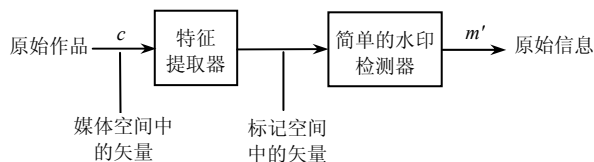


图 3.8 利用标记空间概念的水印检测器框图

通常设计水印嵌入器时并没有明确使用标记空间, 但实际上是可以使用的。如图 3.9 所示, 这样的嵌入器将是一个分为三步的过程。第一步与水印检测器中的提取过程是相同的, 它将未嵌入水印的作品映射成标记空间中的一个点。第二步是在标记空间中选择 一个和提取标记接近的新矢量, 并希望它能够被检测出所包含的理想水印。人们将这个新矢量和原始提取标记之间的差异称为**附加标记**。第二步可以想象成一个简单的水印嵌

入器在标记空间内进行处理。第三步是提取过程的逆过程，它将新矢量映射回媒体空间从而获得含水印作品。这里，人们的目的是为了找到一件作品，由这件作品可得到这个新矢量并作为自己本身的提取标记。如果标记空间和媒体空间具有相同维数，那么这种投影就可以以某种直接的方式进行。然而，如果标记空间的维数比媒体空间小，那么标记空间中的每一点必须对应媒体空间中的很多点。这样就会存在许多件这样的作品，根据它们都能得到这个新矢量并作为提取标记。在理想情况下，人们想要选择一个在视觉上同原始作品最接近的。在实际中，人们通常采用一个近似算法，它虽然不能提供最接近的作品，但是不管怎样它能给出一个比较接近的作品。

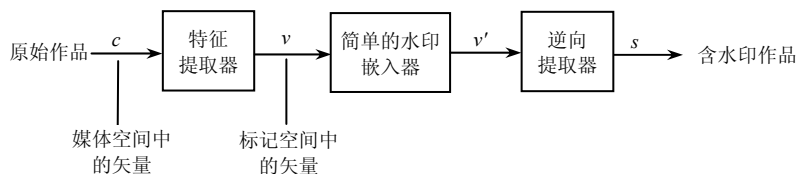


图 3.9 利用标记空间概念的水印嵌入器框图

在图 3.8 和图 3.9 的系统中，设计提取函数的第一个目的是为了减少嵌入和检测的复杂度。第二个目的是为了将未嵌入水印作品的分布、可接受逼真度的区域，以及失真分布进行简化，这样就可设计简单的水印算法。例如，通过将每组采样值进行平均，可以获得一个使不含水印作品的分布更接近于高斯分布（根据中心极限定理）的标记空间。又如，通过频域变换并根据由感知特性决定的常数来进行比例缩放，人们可获得一个使可接受逼真度的区域更接近于球形的标记空间。再如，通过对几何上和时间上的失真进行补偿，人们可以获得一个使失真分布不具备多种形式的标记空间。

3.4 数字水印技术的应用和性能评价

一般来说，如果一段元数据与一项著作的附加信息有关联的话，这段元数据便可作为水印嵌入。当然，将信息与著作联系起来也有很多其他的方法，比如将其置于数字文件头中，将其编码为条形码的形式置于图像中，或者在一段音频之前宣布信息内容作为其介绍。问题在于：什么场合下水印是一种更好的选择？水印能够解决这些简单技术所不能解决的什么问题？实际上，水印技术与其他技术的差别主要体现在以下三个重要方面：

（1）水印是不可感知的。与条形码不同，它不会影响图像的美感。

（2）水印同其所嵌入的作品不可分离。与文件头不同，在作品被展示或者被转变为其他文件格式时，水印不会被清除。

（3）水印能够与作品经历同样的变换。这意味着在某些情况下可以通过观察最终水印而获得作品所经过的变换的一些信息。由于这些性质，水印在一些应用场合下十分有用。一个水印处理系统的性能评价可以基于某些特性指标。例如，**鲁棒性**描述了水印经历常见的信号处理操作而继续存留的能力，**保真度**（Fidelity）描述了水印的不可感知性等。这些特性之间的相对重要性取决于系统设计所适用的具体场合。例如，从通过模拟信道广播的作品中检测水印时，水印必须具有极强的鲁棒性以抵抗信道性能恶化造成的干扰。当然，如果能保证作品在水印嵌入与检测前后不会遭到更改，那么水印的鲁棒性无关紧要。

本节首先介绍几种可以采用水印技术的应用场合，并探究水印同其他技术相比所具有的优势。然后，介绍水印系统的数种特性，并探讨在不同的应用场合下它们之间的相

对重要性及其解释。最后，介绍数字水印系统的评价问题。

3.4.1 数字水印技术的应用

数字水印技术的应用极为广泛，主要有以下七种应用领域^[33]：广播监控、所有者识别、所有权验证、交易跟踪、真伪鉴别、复制控制以及设备控制。下面简要介绍每一种应用，分析问题的特征以及水印作为其解决方案的理由。详细应用介绍见第9章。

1. 广播监控

广告商希望他们从广播商处买到的广告时段能够按时全部播放，广播者则希望从广告商处获得广告收入。为了实现广播监控，可雇用监控人员对所播出的内容直接进行监视和监听，但这种方法不但花费昂贵而且容易出错。或者用动态监控系统将识别信息置于广播信号之外的区域，如视频信号的**垂直空白间隔**（Vertical Blanking Interval, VBI），但是该方法涉及兼容性问题。水印技术可以对识别信息进行编码，是替代动态监控技术的一个好方法。它利用自身嵌入在内容之中的特点，无须利用广播信号的某些特殊片段，因而能够完全兼容于所安装的模拟或数字的广播基础设备。

2. 所有者识别

文本版权声明用于作品所有者识别具有一些局限。首先，在复制时这些声明很容易被去除，有时甚至不是故意为之。例如，一位教授对一本书的某几页进行复制时，很可能会忽略复印主题页上的版权声明。另一个问题是它可能会占据一部分图像空间，破坏原图像的美感且易被剪去除去。由于水印既不可见，也同其嵌入的作品不可分离，故水印比文本声明更利于所有者识别。若用户拥有水印检测器，他们就能够识别出含水印作品所有者，即使用能够将文本版权声明除去的方法来改动它，水印也依然能够被检测到。

3. 所有权验证

除了对版权所有者信息进行识别外，利用水印技术对其进行验证也是令人关注的一项应用。传统的文本声明极易被篡改和伪造，无法用来解决该问题。针对此问题的一个解决办法是建立一个中央资料库，对数字产品的复制进行注册，但人们可能会因费用高而打消注册念头。为了省去注册费用，人们可以使用水印来保护版权，而且为了使所有权验证达到一定安全级别，可能需要限制检测器的发放。如果对手没有检测器，清除水印则是相当困难的。然而，即使水印不能被清除，攻击者也可以使用自己的水印系统，让人觉得数字产品里好像也具有攻击者的水印。因此，人们无须通过所嵌入的水印信息直接证明版权，而是要设法证明一幅图像从另一幅得来这个事实。这种系统能够间接证明有争议的这幅图像更有可能为版权所有者所有而不是攻击者所有，因为版权所有者拥有创作出含水印图像的原始图像。这种证明方式类似于版权所有者可以拿出底片，而攻击者却只能够伪造受争议图像的底片，而不可能伪造出原始图像的底片来通过测试。

4. 交易跟踪

利用水印可记录作品的某个复制所经历的一个或多个交易。例如，水印可记录作品的每个合法销售和发行的复制的接收者。作品的所有者或创作者可在不同复制中加入不同水印。若作品被滥用（透露给新闻界或非法传播），所有者可以找出那个应该负责的人。

5. 内容真伪鉴别

如今以难以察觉的方式对数字作品进行篡改已经变得越来越容易。消息真伪鉴别问

题在密码学中已有比较成熟的研究。**数字签名**是最常用的加密方法，它实际上是加密的消息概要。如果将经过篡改的消息同原始签名相对照，便会发现签名不符，说明消息被篡改过。这些签名均为源数据，须与它们所要验证的作品一同传送。一旦签名遗失，作品便无法再进行真伪鉴别。使用水印技术将签名嵌入作品中可能是一种比较好的解决方法。人们将这种被嵌入的签名称作真伪鉴别印记。如果极微小改动就能造成真伪鉴别印记失效，这种印记便可称作“脆弱水印”。

6. 复制控制

前面所述的绝大多数水印都只能在不合法行为发生之后起作用。例如，广播监控系统只能够在广播商没有播出客户付费广告的情况下被认定为不诚实，而交易跟踪系统也只能够在对手散发非法复制之后被识别出身份。显然，最好能够制止非法行为的发生。在复制控制的应用中，人们致力于防止他人对受版权保护的内容进行非法复制。防止非法复制的第一道防线就是加密。使用特定密钥对作品加密后，可以使没有此密钥的人完全无法使用该作品。然后可以将此密钥以难以复制或分发的方式提供给合法用户。但是，人们通常希望媒体数据可以被观赏，却不希望它被人复制。这时人们可以将水印嵌入内容中，与内容一同播放。如果每个录制设备都装有一个水印检测器，设备就能够在输入端检测到“禁止复制”水印的时候禁用复制操作。

7. 设备控制

复制控制实际上属于更大范围的一个应用——设备控制的范畴。设备控制是指设备能在检测到内容中的水印时作出反应。例如，Digimarc 的“媒体桥”系统可将水印嵌入到经印刷、发售的图像中，如杂志广告、包裹、票据等。若这幅图像被数字摄像机重新拍照，那么 PC 上的“媒体桥”软件和识别器便会设法打开一个指向相关网站的链接。

3.4.2 数字水印技术的特性

下面重点介绍针对不同应用，水印系统应具备的 10 种重要特性。每种特性的相对重要性取决于应用要求和水印所起的作用，甚至对水印特性的解释也会随应用场合变化。与嵌入有关的特性包括^[33]：**有效性**（Effectiveness）、**逼真度**（Fidelity）和**容量**（Payload）。与检测有关的特性包括：**盲检测与明检测**（Blind and Informed Detection）、**虚检行为**（False Positive Behavior）和**鲁棒性**（Robustness）。**安全性**（Security）和**密钥**（Secret Keys）紧密相关，因为密钥的使用总是评估水印方案安全特性不可分割的一部分。

1. 嵌入有效性

如果把一件作品输入水印检测器得到一个肯定结果，人们就可以将这件作品定义为含水印作品。基于此定义，水印系统的有效性指嵌入器的输出含有水印的概率。换言之，有效性指在嵌入过程之后马上检测得到肯定结果的概率。在一些情况下，水印系统的有效性可以通过分析确定，也可以根据在大型测试图像集合中嵌入水印的实际结果确定。只要集合中的图像数目足够大且与应用场合下的图像分布类似，输出图像中检测出水印的百分比就可以近似为有效性的概率。

2. 逼真度（透明性、不可感知性）

一般来说，水印系统的逼真度指原始作品与其嵌入水印版本之间的感官相似度。但如果含水印作品在被人们观赏之前，在传输过程中质量有所退化，那么应该使用另一种逼真度定义。人们可以将其定义为在消费者能同时得到含水印作品和不含水印作品的情

况下，这两件作品之间的感官相似度。在使用 NTSC 广播标准传输含水印视频或者使用 AM 广播传输音频时，由于广播质量相对较差，经过信道质量退化后的原始作品与其含水印版本之间的差异几乎无法让人察觉。但在 HDTV 和 DVD 的视频和音频中，信号质量非常高，则需要高逼真度的含水印作品。

3. 数据容量

数据容量指在单位时间或一幅作品中能嵌入水印的比特数。对一幅照片而言，数据容量指嵌入此幅图像中的比特数。对音频而言，数据容量即指在一秒钟的传输过程中所嵌入的比特数。对视频而言，数据容量既可指每一帧中嵌入的比特数，也可指每一秒内嵌入的比特数。一个以 N 比特编码的水印称作 N -比特水印。这样的系统可以用来嵌入 2^N 个不同的消息。许多应用场合要求检测器能执行两重功能。首先确定水印是否存在，如果存在，则继续确定被编码的是 2^N 个消息中的哪一个。这种检测器有 2^N+1 个可能的输出值： 2^N 个消息和“不存在水印”。

4. 盲检测与明检测

人们将需要原始不含水印的复制参与的检测器称作明检测器。这个名称也可指那些只需要少量原始作品的遗留信息而不需要整件原始作品参与的检测器。而人们把那些不需要原始作品任何信息的检测器称作盲检测器。水印系统使用盲检测器还是明检测器决定了它是否适合某一项具体应用。明检测器只能够用于那些可以得到原始作品的场合。

5. 虚检概率

虚检指在实际不含水印的作品中检测到水印的情况。关于这个概率存在两种定义，区别在于作为随机变量的是水印还是作品。在第一种定义下，虚检概率指在给定一件作品和随机选定的多个水印的情况下，检测器报告作品中发现水印的概率。在第二种定义下，虚检概率指在给定一个水印和随机选定的多个作品的情况下，检测器报告作品中发现水印的概率。在大多数应用中，人们对第二种定义下的虚检概率更感兴趣。但在少数应用中，第一种定义也同样重要，例如在交易跟踪的场合，在给定作品的情况下检测一个随机水印，常会发生虚假的盗版指控。

6. 鲁棒性

鲁棒性指在经过常规信号处理操作后能够检测出水印的能力。针对图像的常规操作包括空间滤波、有损压缩、打印与复印、几何变形（旋转、平移、缩放及其他）等。在某些情况下，鲁棒性毫无用处甚至被极力避免，如水印研究的另一个重要分支就是脆弱水印，它具有和鲁棒性相反的特点。例如，用于真伪鉴别的水印就应该是脆弱的，即对图像做任何信号处理操作都会将水印破坏掉。在另一类极端应用中，水印必须对任何不至于破坏含水印作品的畸变都具有鲁棒性。

7. 安全性

安全性表现为水印能够抵抗恶意攻击的能力。恶意攻击指任何意在破坏水印功用的行为。攻击类型可归纳为三大类：非授权去除、非授权嵌入和非授权检测。非授权去除和非授权嵌入会改动含水印作品，因而可看作主动攻击；而非授权检测不会改动含水印作品，可看作被动攻击。非授权去除是指通过攻击可以使作品中的水印无法检测。非授权嵌入也指伪造，即在作品中嵌入本不该含有的非法水印信息。非授权检测，可以按严重程度分为三个级别：最严重的级别为对手检测并破译了嵌入的消息；次严重的攻击为对手

检测出了水印，并辨认出了每一点印记，但却不能破译这些印记的含义；非严重的攻击为对手可以确定水印的存在，但却不能够对消息进行破译，也无法分辨出嵌入点。

8. 密码与水印密钥

在现代加密算法中，安全性只取决于密钥安全性，而不是整个算法安全性。人们希望水印算法也具有同样的标准。理想情况下，如果密钥未知，即使水印算法已知，也不可能检测出作品中是否有水印。甚至在部分密钥被对手得知也不可能在完好保持含水印作品感官质量的前提下成功去除水印。由于在嵌入和检测过程中使用的密钥与密码术中的密钥所提供的安全性不同，人们经常在水印系统中使用两种密钥。消息编码时使用一个密钥，嵌入过程则使用另一个密钥。为区分两种密钥，分别称为生成密钥和嵌入密钥。

9. 内容修改与多重水印

当水印被嵌入到作品中时，水印的传送者可能会关心水印的修改问题。在一些应用场合不希望水印能够被轻易修改，但在另一些场合修改水印则是必须的。在复制控制中，广播内容会被标明“一次复制”，经过录制后，则被标记为“禁止再复制”。在一件作品中嵌入多重水印的场合是交易跟踪领域。内容在被最终用户获得之前，往往要通过多个中间商进行传播。复制标记上首先包括版权所有者的水印。之后作品可能会分发到一些音乐网站上，每份作品的复制都可能会嵌入唯一的水印来标识每个分发者的信息。最后，每个网站都可能会在每件作品中嵌入唯一的水印用来标识对应的购买者。

10. 耗费

对水印嵌入器和检测器的部署作经济考虑是件十分复杂的事情，它取决于所涉及的模式。从技术观点看，两个主要问题是水印嵌入和检测过程的速度以及需要用到的嵌入器和检测器的数目。其他一些问题还包括嵌入器和检测器作为特定用途的硬件设备实现还是作为软件应用程序实现，或者是作为一个插件实现。

3.4.3 数字水印系统的评价问题

研究数字水印系统的部门和学者都需要对不同水印算法进行评价和比较的方法，对水印应用感兴趣的部门也需要通过评价找到最合适的水印系统。而致力于开发新的水印系统的部门也需要一些改进程度的检验标准。这些标准可以用来对各种特性进行优化。

1. “最佳”的含义

在评价一个水印系统与算法之前，需要弄清一个系统究竟比另一个系统好在何处，怎样才算是最佳性能。如果仅对某个特定的水印应用感兴趣，则评价标准完全取决于具体的应用场合。例如，要评价一个复制控制中的视频水印系统，就需要测试其对微小旋转的鲁棒性，因为微小旋转是一种可能遇到的视频攻击。然而，这个鲁棒性测试在广播监控应用场合中却毫无用处，因为在该场合中人们根本不关心其抵抗主动攻击的问题。

如果人们想要对一个新的水印系统的优点进行测试，选择评价标准时就会有更多的灵活性。一般来说，在相同条件下，如果新系统在某一特性上的性能有所改进，那么就可以说新系统很有价值，至少在此项特性特别重要的应用场合。当然，在某项特性上性能的改进经常要以其他方面的性能降低为代价。例如，假定在一个数据容量为 20 比特的水印系统的基础上，将其数据容量扩充至 30 比特，并保持其虚检概率、鲁棒性和逼真度不变。显然，新系统因为水印数据容量的增大而优于旧系统。但在一些其他应用中，人

们可利用这种较高的数据容量换取更低的虚检概率。比如将嵌入器改为只承载 20 比特的消息，而将剩下的 10 比特作为“校验和”信息随同原消息一起嵌入。检测器随后会抽取全部 30 比特信息，以检查“校验和”是否正确；如果检验结果失败，则报告水印并不存在。这样做实际上得到了一个虚检概率降低了 1024 倍的 20 比特水印系统。

许多水印检测器都具有检测门限参数，人们可降低这个参数，将虚检概率指标的改善转变为鲁棒性的改善。许多嵌入器都采用了嵌入强度参数，利用它可在鲁棒性和逼真度（不可见性）之间进行折中。如果可以验证新的系统比旧的系统具有更大的数据容量，便可声称它适合更为广泛的应用。

2. 测试基准 (Benchmarking)

基准 (Benchmark) 是一个标准，任何特例都依据这个标准来测量。基准测试是指一个在全球范围内，持续定义、比较、配置和评估最优实践的构造和分析过程。水印系统的评测基准用来比较不同水印嵌入算法的优劣（主要指抵抗各种攻击的能力），并给出相应的得分和图表数据。水印评测基准的输入参数包括：攻击类型、信噪比、密钥数、性能评价准则、消息、载体、生成/嵌入/检测/解码算法。输出结果包括嵌入时间、正确水印的检测/解码的平均时间、错误水印的检测/解码的平均时间等。测试标准一旦确定，就可把注意力转向测试程序的开发。在评估某个特定应用的水印系统或算法时，对每条特性都应列出其评估标准的最低指标要求，然后开发一套测试程序以检查系统是否满足这些要求。对给定系统进行测试时，嵌入器与检测器必须基于相同的参数设置（如固定的嵌入强度和检测门限）。因水印研究的需要，人们对开发一套通用水印测试基准很感兴趣。研究者可以使用这种测试基准对所提出的水印系统打出一个单独、量化的“分数”。这个分数可用来和其他类似的被测试系统进行比较。例如，我们可以提出用于图像水印系统的测试基准如下：指定必须嵌入的数据容量比特数（比如 80）以及必须达到的逼真度（通过一种特定的感知模型来测定，比如 JND（刚好可察觉失真））。在这两种特性参数保持一定的情况下，使用 StirMark 程序（将在下面介绍）对多种失真的鲁棒性进行测试。这个程序可以使图像具有对感官质量影响很小的失真，但这些失真却能够使大多数水印无法被检测到。水印系统在测试中得到的各种性能指标被综合为一个单独的分数。然而，这种测试基准只能针对某些数字水印系统，故只能评价数量有限的水印技术。实际上，有些水印系统具有非常小的数据容量（如 8 比特或更少），非常好的鲁棒性，非常小的虚检概率，或非常小的耗费。这类系统往往采用了无法扩展到 80 比特的编码。而且许多对安全要求并不严格的应用场合也并不一定需要 StirMark 提供的测试。另外，对于攻击者很可能拥有水印检测器的一些应用场合，StirMark 并未包含其安全性的全面测试。因此，可以确信，不可能存在适用于所有水印系统和应用场合的测试基准。目前最著名的五种水印评测基准是：unZign、StirMark、Certimark、Checkmark 和 Optimark。

1) unZign

unZign 是一个测试水印鲁棒性的工具。unZign 有 1.1 版本和 1.2 版本。在 1.1 版本中，unZign 使用很细微的图像平移引入像素抖动 (Jittering)。此工具的使用效果依赖于待检验的水印技术，不过似乎对去除或破坏水印很有效。然而，在去除水印的同时，1.1 版的 unZign 经常带来不可接受的图像降质。1.2 版有所改进，图像降质有所减小，但同时破坏水印的能力也同样削弱。unZign 软件包括两个文件：一个为可执行文件 unZign12.exe（以版本为 1.2 的软件为例），另一个为动态连接库 cygwin.dll 文件。unZign 的使用非常简单，在 Windows 9x/Windows NT 的命令行下输入 unZign12 [signed_

file.jpg unsigned_file.jpg]就可对含水印图像进行水印去除操作。如果没有给出输入和输出文件, unZign 将给出版本信息和使用说明。需要注意的是, unZign 只支持 JPEG 格式的图像文件。

2) StirMark

StirMark 是 Petitcolas 等在英国剑桥大学攻读博士期间开发的首个用于数字图像水印算法简单鲁棒性测试的通用工具, 其第一版发表于 1997 年 11 月。随后几个改进的版本 1.0、2.2、2.2b、2.3 (针对攻击方法和测试结果的描述) 相继发表。在 1999 年 1 月, 由于公正评测水印系统的迫切需要使得第一个水印评测基准终于以 StirMark 3.1 版正式发布。从 3.0、3.1 版到现在的 4.0 版和 4.1 版, StirMark 在水印界获得极大的关注, 它已成为目前最为广泛使用的用于水印攻击的基准测试工具。StirMark 对给定的一幅含水印图像进行测试, 就能生成许多修改后的图像, 以此用来验证嵌入的水印是否仍能被检测到。StirMark 还能结合不同的检测结果计算出一个在 0~1 之间的综合分数。

StirMark 引入了全局几何形变和局部形变。全局几何形变包括: 旋转、比例缩放、方向比例变化、平移和剪切等, 这些都属于广义仿射变换类。行/列去除和裁剪/平移也整合在 StirMark 中。近来的一些水印方法, 采用了特殊的同步技术, 在这些攻击下大多都能幸存。对全局几何形变的鲁棒性依赖于变换不变域的使用、附加的模板或水印本身的自相关函数 (ACF)。如果说对全局仿射变换的鲁棒性是一个已解决的问题, 那么整合在 StirMark 中的局部随机扭曲几乎对所有的技术仍是一个公开问题。这个所谓的随机扭曲利用了这样一个事实, 即人类视觉系统对移位和局部仿射修正是不敏感的。因此, 像素在没有显著的视觉形变下被局部移位、缩放和旋转。对各种水印方案用此软件进行测试, 可得到各种方法的对比结果。此软件支持多个平台: Linux、Windows 9x/Windows NT、Macintosh 等。各个平台各个版本的 StirMark 软件都可直接从网上获得, 对于 StirMark 4.0, 下载网址为: http://www.petitcolas.net/fabien/software/StirMarkBenchmark_4_0_129.zip。

下面以 Windows 下的 StirMark 3.1 为例, 简要介绍它的使用。StirMark 3.1 和 unZign 一样也是一个命令行软件。这么做的好处是可编写成批处理程序, 便于成批测试。StirMark 没有安装程序, 直接将从网上下载的压缩文件用 Pkunzip 或 Winzip 解压缩到一个目录下即可。解压以后, StirMark.exe 文件就是程序的可执行文件。只要在命令行下按如下格式输入命令即可启动 StirMark 软件进行测试: Stirmark [options] [<input file> [<output file>]]。其中, options 为可选参数, 可能的参数约有 18 种。输入和输出文件是可选项, 如果输入和输出文件名没有出现在命令行中, 则程序在运行后会要求你输入文件名称。StirMark 能够读/写的文件包括 PGM、PPM、JPEG 三种文件, 在 Windows 下分别以 .pgm, .ppm, .jpg 为后缀。StirMark 能够根据文件后缀自动分辨文件类型。下面给出一些具体的例子: ① Stirmark-Tbase test.ppm, 对 test.ppm 进行基准测试, 并将测试结果存在以 base 开头的各文件中; ② Stirmark-S34-Tbase test.ppm, 对 test.ppm 进行 34 号基准测试, 并将结果存在以 base 开头的各文件中; ③ Stirmark-PSNR original.ppm modified.jpg, 计算 modified.jpg 相对 original.ppm 的峰值信噪比 PSNR; ④ Stirmark-NOJPEG-PS+sample.prm sample.ppm sample_S.ppm, Stirmark-NOJPEG-PL+sample.prm sample.ppm sample_SL.ppm。第一个命令对 sample.ppm 图像进行几何失真, 并且失真后不进行 JPEG 压缩。同时, 将各参数存入 sample.prm 文件中。第二个命令将 sample.prm 中的参数调入, 并使用这个参数对 sample.ppm 图像进行几何失真。⑤ Stirmark-i0-o0-d0 test.ppm output.jpg, 未作任何处理, 只是将 test.ppm 文件转为 output.jpg 文件。

3) Checkmark

Checkmark 是由 Pereira 开发的一种基准测试工具，是在 UNIX 或 Windows 平台下运行于 Matlab 上的用于数字水印技术的一组基准套件。第二代水印测试软件 Checkmark 具有 StirMark 全部的攻击类型，Checkmark 包含了一些未在 StirMark 中提出的攻击。而且，它还考虑了水印应用，这意味着从单个攻击得出的分数将根据它们对于一个给定的水印用途的重要性进行加权；因此它提供了一种更好地评估水印技术的有效工具。它给出重要的新的攻击类：① 小波压缩（基于 JASPER 的 JPEG2000）；② 投影变换、视频失真建模、扭曲；③ 复制攻击；④ 模板去除攻击；⑤ 反噪声、伴随知觉重调制的反噪声；⑥ 非线性行去除；⑦ 拼贴攻击、下/上采样、抖动、门限化。这些都是水印嵌入技术中面临的待解决的新问题。Checkmark 最初的 1.0 版是在 2001 年 6 月 10 发布的，后来又发布了 1.0.2、1.0.4、1.0.5 版，最新的 Checkmark 是在 2001 年 12 月 14 日发布的 1.2 版，已支持彩色图像，在线 FAQ（常见问题解答），并更新了在线结果。与 StirMark 相比，Checkmark 添加了新的质量测量方法——加权 PSNR 和 Watson 测量方法；以灵活的 XML 格式输出和生成 HTML 结果表格；容易将 Matlab 上的单个攻击用于测试等。网站为：<http://cvml.unige.ch/ResearchProjects/Watermarking/Checkmark/>。

4) Optimark

Optimark 是用于静态图像水印算法的一个基准测试工具，它由希腊 Thessaloniki 的 Aristotle 大学信息学系的人工智能和信息分析实验室开发。目前，Optimark 最新的版本是于 2002 年 1 月 29 日更新的 1.0 版。与 StirMark 和 Checkmark 不同的是，Optimark 具有图形界面；它能利用不同的水印密钥和信息，使用多重测试进行检测/解码性能评估。Optimark 针对水印检测器给出的不同结果（浮点结果或二值结果），相应给出不同的性能测量方法的评估。此外，Optimark 还提供了对解码性能的测量方法、平均嵌入和检测时间、算法有效载荷以及某一攻击和某一性能标准的算法崩溃极限的评估。使用用户在选定的攻击和图像上定义的权值后，Optimark 能给出多重等级的结果，并且用户还可以选择自定义和事先设置基准部分。表 3.1 比较了 StirMark、Checkmark 和 Optimark 支持的主要攻击类型。感兴趣的读者可访问网页 <http://poseidon.csd.auth.gr/optimark/>。

表 3.1 各种基准软件所支持的攻击类型

攻击类型	StirMark	Checkmark	Optimark
裁剪	✓	✓	✓
翻转	✓	✓	✓
旋转	✓	✓	✓
旋转-尺寸缩放	✓	✓	✓
FMLR*	✓	✓	
锐化	✓	✓	✓
高斯滤波	✓	✓	✓
中值滤波	✓	✓	✓
随机扭曲	✓	✓	*
线性变换	✓	✓	✓
方向比例	✓	✓	
缩放改变	✓	✓	✓
行移除	✓	✓	✓

续表

攻击类型	StirMark	Checkmark	Optimark
颜色降质	√	√	
JPEG 压缩	√	√	√
小波压缩		√	
投影变换		√	
扭曲		√	
模板移除		√	
非线性行移除		√	
拼贴		√	

注 *表示支持旋转+自动裁剪和旋转+自动裁剪+自动缩放。

5) Certimark

Certimark (Certification of Watermarking Techniques) 是用于视觉内容水印嵌入和水印算法鉴定的基准软件。该项目是欧盟的项目，于 2000 年 5 月立项，2002 年 7 月完成。该项目的目的是：（1）设计和开发一套完整的用于水印技术（包括静态图像和视频）评测的基准软件；（2）研究高效的有实力的水印算法。

该软件的模块如图 3.10 所示，各模块解释如下：

- （1）图像源，即待嵌入水印的原始图像及其有关文件格式、大小等参数；
- （2）待测系统（System Under Test, SUT）的水印嵌入器，这里包括影响嵌入的各种参数，如密钥、嵌入信息量等；
- （3）攻击模块，该模块用来模拟各种无意的和恶意的攻击；
- （4）待测系统（System Under Test, SUT）的水印检测器，这里包括影响提取的各种参数；
- （5）比较模块与视觉质量；
 - ① 比较模块，比较提取的水印信息与原始水印信息；
 - ② 与比较模块平行的主观/客观质量评价模块；
- （6）报表记录器模块（Report Writer Module），该模块生成评测报告，以概括方式提供给用户；
- （7）鉴定模块，该模块考虑不同的准则和应用场合，证明所测算法在特定应用下的有效性。

感兴趣的读者可参见网页 <http://www.certimark.org> 以及“Certimark 的结构和未来展望”一文。

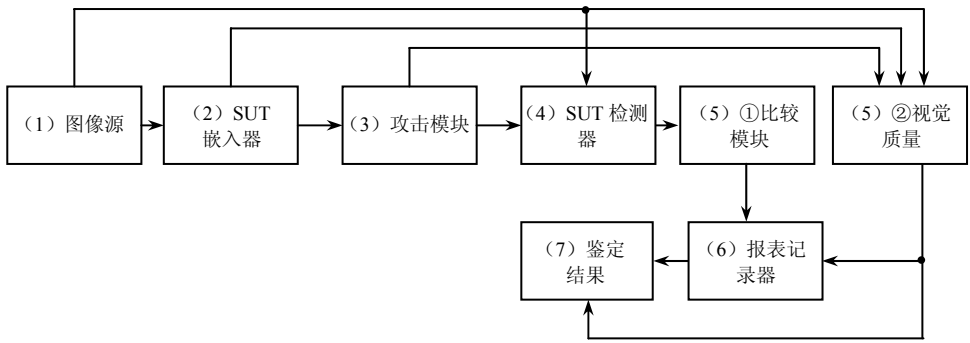


图 3.10 Certimark 软件框图

3. 测试范围 (Scope of Testing)

一般来说, 水印系统或算法应使用与具体应用分布类似的大量作品进行测试。例如, 我们不能指望某个适合于 Lena 图像的算法也同样适合于 X 光图像、卫星图像或者动画。如果对系统的测试并不针对某一项具体应用, 那么用于测试的作品应能够代表一类应用。

测试集合的尺寸与典型性这两个问题对于虚检概率的测试尤为重要。可以设想一个检测器要在大量作品中搜索一个特定的水印模式。若该系统要求具有 10^{-6} 的虚检概率, 则意味着检测器对每一百万件作品进行检测, 发生虚检的次数最多不能超过一次。为了验证这种性能, 人们可能需要给检测器提供数以百万计的作品。显然在很多情况下, 这种测试是不可行的, 因此人们必须设计一些符合统计模型的测试过程用于预测虚检概率。虽然这些测试不需要上百万件作品, 但数量太少的话也容易导致令人误解的结果。

3.5 数字图像水印技术

近 20 年来, 随着多媒体技术的飞速发展, 数字图像已成为主要的网络交互对象之一, 图像交互在军事系统、政府机构、金融系统和医疗保健等重要机构中得到广泛应用。图像的安全问题受到了人们的普遍重视。如何保证数字图像的安全交互, 防止图像被第三方截获、恶意篡改、非法复制和破坏图像版权变得非常重要。图像的安全性需求主要包含以下方面。① 保密性: 禁止第三方非法获取图像信息; ② 完整性: 防止图像信息在存储或传输过程中被破坏、丢失和篡改; ③ 版权保护: 通过有效的手段能够辨别图像的真伪, 并提供有效的法律依据。图像安全技术包括图像加密和图像水印。图像加密用于保证图像的保密性。图像水印用于图像的完整性验证和版权保护, 是图像加密技术的重要补充。图像水印是在不影响图像正常使用的情况下, 在图像的空间域(变换域), 按某种变换规则, 改变像素(变换域系数)的位置或值, 在数字图像中嵌入一些版权标记信息(水印), 以达到保护图像产品版权或完整性验证的目的。下面首先介绍数字图像水印系统的基本要求, 然后介绍数字图像水印系统的基本模型和算法分类, 接着介绍数字图像水印系统的关键技术, 然后介绍数字图像水印算法的评价指标, 最后分别介绍典型的鲁棒数字图像水印算法和脆弱数字图像水印算法。

3.5.1 数字图像水印系统的基本要求

数字图像水印技术作为信息隐藏学的一个分支, 是通过一定的算法将一些标志信息嵌入到图像内容当中, 但不影响原图像内容的价值和使用, 并且不能被人的视觉系统觉察或注意到, 只有通过专用的检测器或阅读器才能提取。为了更好地实现数字图像的真伪鉴别和版权保护等目的, 数字图像水印算法必须具有以下特性^[33]。

1. 鲁棒性 (Robustness)

指图像水印算法抵抗常见图像处理操作的能力, 也就是说含水印图像经历无意修改而保留水印信息的能力。一般来说, 水印应当具有对噪声、平滑、增强、有损压缩、平移、旋转、缩放和裁剪具有鲁棒性。但是, 对于内容认证应用场合, 必须要求水印具有脆弱性 (Fragility)。

2. 透明性 (Transparency)

透明性也称不可见性, 即水印的存在不应明显干扰被保护的图像数据。换句话说,

数字水印的嵌入不应使原始图像数据发生可感知的改变,也不能使得被保护图像数据在质量上发生可以感觉到的失真。通常需要利用人类视觉系统特性来进行水印处理,从而使含水印图像没有明显的主观降质现象,而嵌入的水印却无法为人地看见。需要指出,对于广告和版权通知应用场合,则通常会利用可见水印技术来起到广告宣传或警示作用,即使这样,人们还是希望对图像载体的质量影响越小越好。

3. 安全性 (Security)

指水印算法抵抗恶意攻击的能力,即它必须能够承受一定程度的人为攻击,而使水印信息不会被删除、破坏或窃取。应该保证非授权用户无法检测或破坏水印,即使在水印算法或相关知识公开的情况下。安全性中还包含不可检测性 (Undetectability),指隐蔽载体与原始载体具有一致的特性 (如具有一致的统计噪声分布等),以便使非法拦截者无法判断是否有隐蔽信息。

4. 数据容量 (Data Capacity)

水印应该能够包含相当的数据容量,以满足多样化的需要。

5. 盲检测和自恢复性

盲检测是指水印的检测和提取不需要原始图像的参与。而自恢复性的含义如下:含水印图像经过一些操作或变换后,可能会产生较大的失真或破坏。如果只从留下的片段数据仍能恢复水印信号,而且恢复过程不需要原始图像参与,这就是所谓的自恢复性。这个要求显然比盲检测要求稍高一些。

6. 确定性 (Unambiguous) 和可证明性

水印所携带的信息能够被唯一确定地鉴别。显然,只有保证足够的数据容量才能确保确定性。水印必须能为信息产品的归属提供完全和可靠的证据。

7. 篡改定位能力

这一点主要针对数字水印的图像内容认证场合。是指识别作品被改变的时间和区域,并证明作品的其余部分未发生改变的能力。

一般而言,上述这些特性要求之间通常是相互竞争和矛盾的,实际应用中不可能使它们同时达到最佳,只能根据需要在不同特点之间取得折中或者设定可调范围。

3.5.2 数字图像水印系统的基本模型和算法分类

数字图像水印处理过程主要包括水印生成、嵌入和检测三个步骤,而整个水印系统还应包括外界的攻击过程。图 3.11 给出了整个数字图像水印系统的基本模型,虚线表示对应项可以参与操作或可以不参与操作。数字水印生成算法 G 的输入为原始信息 m 、原始图像 c 和水印生成密钥 k_1 ,输出为待嵌入的数字水印 w 。当然,在生成过程中也可以不需要原始图像,甚至也不需要原始信息,直接由密钥(种子)生成水印序列。人们通常采用的水印形式是二进制序列,例如:由伪随机序列发生器产生的伪随机二进制序列信号。有时,数字水印并不通过生成算法生成,而直接给定有意义的图形或图标作为数字水印。在水印嵌入过程中,原始图像 c 、水印 w 以及嵌入密钥 k_2 经过嵌入函数 E 的作用,生成含水印图像 s 。需要注意,尽管图中的生成密钥和嵌入密钥都用虚线表示,但必须确保至少有一个过程使用密钥以保证水印算法的安全性。通常,嵌入函数 E 用插入操作符 \oplus 作用在一组特征集 $F(c)$ 来描述: $F(s) = F(c) \oplus w$ 。根据水印所嵌入的特征集类型的

不同, 数字图像水印算法一般分为空域算法和变换域算法两类(通常还结合了扩频通信理论以及人类视觉系统特性等)。在空域算法中, 水印被嵌入在图像的亮度或色度分量中; 而在变换域算法中, 位于特征集 $F(c)$ 中的特征是图像的变换域系数(例如离散余弦变换、离散傅里叶变换或小波系数等)。一旦生成含水印图像 s , 该图像将在一定的媒介中传输或流通, 一定会受到一些有意或无意的攻击, 从而得到可疑图像 s' 。关于水印检测、提取或解码的含义, 在这里需要作特殊说明。通常把根据检测密钥 k_3 (有时还需要原始图像 c 和原始水印 w) 判断可疑图像 s' 是否存在水印的过程称为水印检测, 而把根据提取密钥 k_3 (有时还需要原始图像 c) 提取可疑图像 s' 中的水印 w' 的过程称为水印提取(或水印恢复)。通常, 在包含水印生成的水印处理系统中, 在提取出水印 w' 后有时还需要根据水印生成密钥 k_1 对应的解码密钥 k_4 恢复所嵌入的信息 m' , 这个过程就是水印解码过程。有时, 提取的水印 w' 没有感知上的含义, 这时需要计算水印 w' 和原始水印 w 之间的相关系数来判断水印的有无, 这个过程相当于先提取后检测。为了描述方便, 通常人们把这三个概念统称水印解码或水印检测。检测或提取密钥 k_3 可以与嵌入密钥 k_2 相同也可以不相同(非对称水印系统); 解码密钥 k_4 也可以与生成密钥相同也可以不相同(公钥水印系统)。在水印提取情况下, 解码函数 D 作用于需要确定版权的可能受损的图像 s' , 并从 $F(s')$ 的特征中恢复水印 $w'=D(F(s'))$ 。同时, 在水印恢复的过程中, 是否要求助于原始图像, 对数字水印的鲁棒性以及安全性等方面都有不同的影响。显然, 盲检测(不需原始图像)的水印方案实用性较好, 而明检测(有原始图参与检测)的水印算法具有较强的鲁棒性。

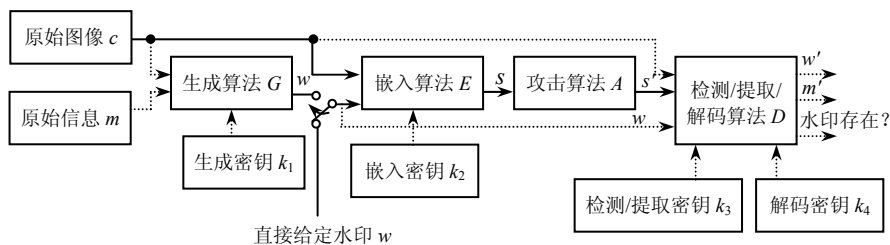


图 3.11 数字图像水印系统模型

图像水印方向已取得丰富的研究成果, 人们已设计出众多的图像水印算法。前面 3.2.3 节已经对数字水印及其技术的分类作了详细说明, 这里再强调一下两种最常见的分类方法。第一, 按照水印的特性, 图像水印技术可分为三类: 鲁棒图像水印技术、半脆弱图像水印技术和脆弱图像水印技术。第二, 按照水印的嵌入域, 图像水印技术可分为三类: 原始域图像水印技术、变换域图像水印技术和压缩域图像水印技术。

3.5.3 数字图像水印系统的关键技术

除了水印攻击外, 数字图像水印系统包括水印生成、水印嵌入和水印检测三个关键技术。下面分别进行简要概括。

1. 数字水印生成技术

数字水印生成是数字水印处理过程的第一步关键步骤。构成水印的序列通常应该具有不可预测的随机性(Unpredictable Randomness)。由于人类视觉系统对纹理具有极高的敏感性, 故水印不应含有纹理。水印应该具有与噪声相同的特性。因此, 目前文献中一般取下述随机序列作为水印嵌入到载体数据中, 具体如下。① 高斯白噪声: 满足均值为

μ 、方差为 σ^2 的正态分布。用得最多的是均值为 0、方差为 1 的高斯白噪声，通常记为 $N(0,1)$ ，这是 Cox 首先提出的一个重要建议。② 伪随机序列：具有类似白噪声的性质，但又具有周期性和规律性，可以人为地加以产生和复制。通常可采用二值的 m-序列、M-序列、混沌序列或其他特殊序列（如勒让德序列）作为水印。有时也可能采用实伪随机序列作为水印。③ 根据有特定含义的原始信息所生成的随机序列：通常选取具有特定意义的字符串或数据段作为原始信息，把每个字符或数据作为产生随机序列的种子，最常见的是伪随机处理（排序、相乘、异或）和扩频两种方式。待嵌入水印序列的元素取值通常有以下几种：① 二值水印，值域可以是单极性的，如 $\Omega=\{0,1\}$ ，也可以是双极性的，如 $\Omega=\{-1, 1\}$ 或 $\Omega=\{-r, r\}$ ，其中 r 为实数。② 三值水印，值域为 $\Omega=\{-1,0,1\}$ 或 $\Omega=\{0, 1, 2\}$ 。③ 整数序列，值域为区间 $\Omega=[-p, p]$ ， $p \in \mathbf{Z}^+$ ，如 CDMA 技术生成的序列、直接给定的可见灰度水印和 VQ（矢量量化）压缩得到的整数索引。④ 实数序列，值域为区间 $\Omega=[-t, t]$ ， $t \in \mathbf{R}^+$ ，如高斯白噪声、实伪随机序列、自适应生成的某些水印序列或经过某些变换的水印序列。⑤ 复数序列，如经过复正交变换的水印序列。考虑到对各种攻击（主要针对几何攻击）的鲁棒性，有时还会考虑采用特殊形状的水印，如六边形水印、圆形或环形水印和自相似水印等。

通常意义上说，数字水印生成过程就是在密钥 k 的控制下由原始版权信息、认证信息、保密信息或其他有关信息 m 生成适合于嵌入到原始载体 c 中的待嵌入水印信号 w 的过程。数字水印生成过程如图 3.12 所示。原始信息有时也称原始水印，其主要类型有如下几种情况：① 文本消息。如 ID 序列号、签名、文本文件或消息。② 声音信号。如语音、音乐或音频信号，但是目前文献中很少提到用声音数据作水印。③ 二值图像。如二值的图片、图章、商标和签名图像。④ 灰度图像。如灰度的商标、图片、照片或图章。⑤ 彩色图像。如彩色的商标、照片或图片等。⑥ 无特定含义（甚至是随机的）的序列。如一维二值序列、一维三值序列、二维二值阵列、实数序列等。关于原始水印序列的长度问题，EBU（European Broadcasting Union）标准和许多文献认为 64 比特的信息足够用来版权识别。因此，为了与载体对象相匹配，往往要对这 64 比特按一定的片率进行扩频处理或者对这 64 比特进行周期延拓，达到所需的长度为止。这样做的好处是可以提高鲁棒性。在实际文献中，也有的学者避开这个长度问题，认为实际嵌入多少比特，长度就为多少。

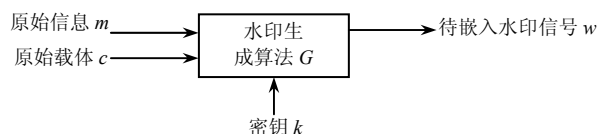


图 3.12 水印生成算法框图

目前，学者们已经提出各种各样的水印生成算法，大致可分为伪随机、扩频、混沌、纠错编码、变换、分解、自适应等生成方法。从各输入项的参与情况来看，数字水印生成方式应该有 7 种情况，可分为以下四类^[33]。

（1）原始信息参与的自适应水印生成方式

这种方式通常需要原始信息、原始载体和密钥的参与，可以用如下函数来表示

$$w = G(m, c, k) \quad (3.9)$$

如果生成过程中不需要密钥，则生成函数可以表示为

$$w = G(m, c) \quad (3.10)$$

需要指出，人们通常把利用原始载体部分或全部信息的生成方法称为自适应生成方

法。这类方法的好处是使鲁棒性和不透明性达到较好折中，有时便于原始载体的自恢复。

(2) 无原始信息参与的自适应水印生成方式

在自适应水印生成过程中，有时不需要原始信息参与，直接根据密钥由原始图像经过一定变换和操作生成待嵌入水印。这时，生成函数可表示为

$$w = G(c, k) \quad (3.11)$$

如果生成过程中不需要密钥，则生成函数可以表示为

$$w = G(c) \quad (3.12)$$

(3) 原始信息参与的非自适应水印生成方式

这种生成模式是文献中用得最多的方式，通常需要密钥参与，其生成函数可表示为

$$w = G(m, k) \quad (3.13)$$

如果生成过程不需要密钥，则生成函数可以表示为

$$w = G(m) \quad (3.14)$$

即直接由原始信息不通过密钥生成水印，如重复放置小水印图变成大水印图用于图像认证的场合，甚至 $w = m$ （即水印直接给定）。

(4) 无原始信息参与的非自适应水印生成方式

这种生成模式也是文献中经常用到的，主要表现为直接利用某个密钥 k 生成伪随机序列和混沌序列的情况，其生成函数可表示为

$$w = G(k) \quad (3.15)$$

下面给出一个水印生成实例。设原始信息为大小 $N \times N$ 的图像，则在水印生成时可采用一个简单的传统混沌系统——**猫映射**（Cat Map），也称 Arnold 变换，其矩阵形式为

$$\begin{pmatrix} x_{i+1} \\ y_{i+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_i \\ y_i \end{pmatrix} \pmod{N} \quad (3.16)$$

其中， (x_i, y_i) 表示图像某像素在第 i 次迭代后的坐标位置。图 3.13 给出了该映射的一个仿真实例，其中图（a）为原始猫图像，图（b）为各像素位置经过 1 次迭代得到的结果图，图（c）为迭代 30 次的结果图。

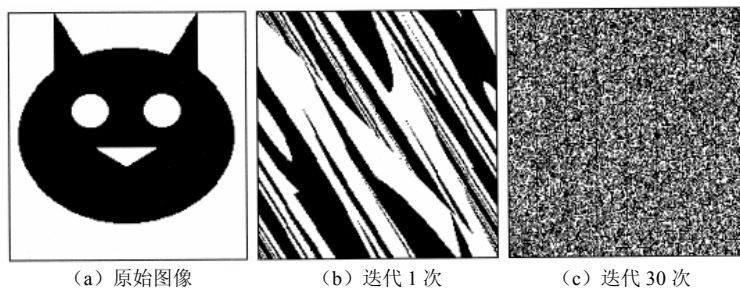


图 3.13 猫映射的仿真实例

2. 数字水印嵌入技术

数字水印系统的第二大关键技术是数字水印嵌入技术。根据所基于的域不同，数字水印嵌入技术主要分为时/空域、变换域和压缩域三大类。时空域算法将水印信息直接嵌入到音频时域采样、图像空间像素和视频数据（按帧或者沿时间轴）等原始载体数据中，即在媒体信号的时间域或空间域上实现水印嵌入。变换域算法将水印信息嵌入到音频、图像、视频、三维目标等原始载体的变换域系数中。压缩域算法广义上是指充分考虑 JPEG、MPEG 和 VQ（矢量量化）技术的结构和特性，将水印嵌入到压缩过程的各种

变量值域中,以提高对相应压缩技术或压缩标准攻击的鲁棒性为目标的嵌入算法。狭义上就是指水印嵌入到 JPEG 位流、MPEG 位流和 VQ 索引流中。变换域水印算法是图像水印技术的主流,因此在水印嵌入前,载体图像需要通过 DCT、DWT 等变换从空域变换到变换域。水印嵌入后,需要通过对应的逆变换实现从频域到空域的变换。水印嵌入算法的流程如图 3.14 所示。

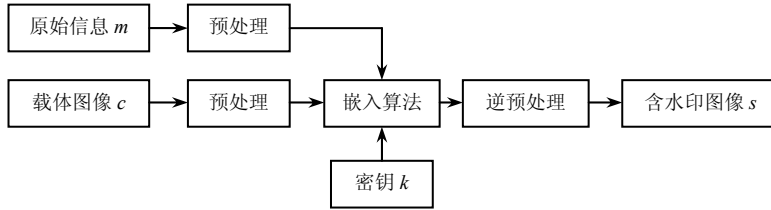


图 3.14 水印嵌入流程

设载体图像可表示为二维信号 $c=\{c_{ij} \mid 0 \leq i < N_1, 0 \leq j < N_2\}$, 通过行扫描可看成一维序列 $c=\{c_i \mid 0 \leq i < N\}$, 长度 $N=N_1 \times N_2$ 。设水印为二维形式 $w=\{w_{ij} \mid 0 \leq i < M_1, 0 \leq j < M_2\}$, 通常 $w_{ij} \in \{-1, 1\}$ 或 $w_{ij} \in \{0, 1\}$, 通过行扫描可看成一维序列 $w=\{w_i \mid 0 \leq i < M\}$, 其长度为 $M=M_1 \times M_2$ 。常用的水印嵌入准则(以空域为例)包括以下几种。

(1) 加法准则

一种最常见的嵌入规则就是加性规则, 不仅适用于空域, 还适用于变换域。通常, 在加性规则中都带有嵌入因子, 以调整所嵌入水印的不可见性和鲁棒性。在空域内, 嵌入公式如下

$$s = c + \alpha w \quad (3.17)$$

其中, $s=\{s_i \mid 0 \leq i < N\}$ 为含水印载体, $c=\{c_i \mid 0 \leq i < N\}$ 和 $w=\{w_i \mid 0 \leq i < N\}$ 分别为原始载体和水印, α 为嵌入因子。基于此准则设计的图像水印算法通常为明检测算法。水印提取时需要载体图像参与。

(2) 乘法准则

在空域中, 乘性规则有两种形式, 一种是将原始载体数据直接乘上水印, 表示如下

$$s = \{s_i \mid s_i = w_i c_i, 0 \leq i < N\} \quad (3.18)$$

另一种形式是在上式的基础上乘上嵌入因子再加上原始载体数据, 表示如下

$$s = \{s_i \mid s_i = c_i + \alpha w_i c_i, 0 \leq i < N\} \quad (3.19)$$

基于此类准则设计的图像水印算法为明检测算法, 水印提取时需要载体图像参与。

(3) 替换准则

替换是一种实现盲提取的重要嵌入手段, 其主要思想是用水印信息或水印信息对应的数值去替换原始载体的数据或特征量, 提取时只需直接提取含水印载体对应的数据或特征量即可。替换方法不仅可用在时域, 也可用在变换域和压缩域中。替换准则中的一种常用准则称为最低有效位 (LSB) 准则。LSB 位平面替换嵌入公式可描述如下

$$s = \left\{ s_{ij} \mid s_{ij} = \begin{cases} c_{ij} + w_{ij} & c_{ij} \text{ 为偶数} \\ c_{ij} - 1 + w_{ij} & c_{ij} \text{ 为奇数} \end{cases}, 0 \leq i < N_1, 0 \leq j < N_2 \right\} \quad (3.20)$$

其中, s 为含水印载体, c 和 w 分别为原始载体和二值水印。实际上, 上式相当于在嵌入之前先将原始载体图像的每个像素的最低有效位清零, 即 LSB 位平面清零, 然后在嵌入时直接加上二值水印即可, 即用二值水印直接替换原有的 LSB 位平面。

(4) 修改统计特征的嵌入准则

利用统计特征是空域嵌入算法中的一类重要嵌入技术。这类方法的主要思想是通过修改原始载体数据使得原始载体的某些统计特征发生变化,检测时只需查看含水印载体的统计特征即可,从而达到盲检测目的。当然,这些统计特征的来源必须受到密钥的控制以保证安全性。在空域嵌入算法中,常用的统计量包括平均值、标准偏差和直方图等。

(5) 基于量化索引调制的嵌入准则

基于量化索引调制思想^[28]的水印嵌入算法的主要目的是为了实现盲检测。其主要思想是根据水印信息的不同将原始载体数据量化到不同的量化区间,而检测时根据数据所属的量化区间来识别水印信息。

(6) 基于关系的嵌入准则

基于关系的嵌入方法的主要目的是为了实现盲检测。在水印嵌入过程中,通过修改载体数据使得水印的不同取值反映了不同的关系,如大小关系、逻辑关系和奇偶性等,从而在检测时根据关系得到相应的水印信息。这种嵌入方式不仅适用于空域,也适用于变换域和压缩域。相邻像素值或变换域系数一般都具有很强的相关性,可通过调整该点数值与其邻域均值关系来嵌入水印。在空域中,常见的方法就是利用像素间的关系、块内或块间的统计特征量的关系或利用邻域像素与中心像素间的关系来嵌入水印。

(7) 交换准则

基于此准则设计的图像水印算法为盲检测算法。例如,我们可以通过一定的规则,选取一对变换域相邻系数 c_i 和 c_j 。如果 $c_i < c_j$ 且待嵌入水印位为 1,则把 c_i 和 c_j 交换位置;如果 $c_i \geq c_j$ 且待嵌入水印位为 0,则交换 c_i 和 c_j 位置。其他情况则保持 c_i 和 c_j 不变。

(8) 基于树结构的嵌入准则

树结构是小波变换中常用的一种结构,这是由小波变换的多分辨率分析方式所决定的对于二维图像的金字塔式小波分解,较粗糙一级的每一个系数对应较精细一级的四个系数,这就隐含地存在一种四叉树结构。这种结构广泛地应用到图像压缩和信息隐藏等领域。在小波变换中,人们常用的两种树为**零树**(Zerotree)和**重要树**(Qualified Significant Wavelet Tree, QSWT)。在 DCT 变换中,也可以提出构造出类似小波变换的零树结构。例如,基于零树结构的嵌入准则可以从三个不同方面考虑:① 将零树内的所有系数根据水印位的不同置成不同极性的同一实数;② 通过改变零树个数的奇偶性来嵌入水印;③ 通过改变零树内元素个数的奇偶性来嵌入水印。

(9) 自适应嵌入准则

自适应嵌入方式是一种重要的嵌入方式。它充分考虑原始载体的局部特征,使得水印嵌入的位置、强度随着局部特征的变化而变化,最终得到不可见性和鲁棒性的一种最佳折中。此外,在图像认证中的一种重要嵌入方式就是自嵌入,也就是水印来源于图像的特征,把特征嵌入到载体本身,以达到篡改定位和自修复的目的,这种方式也是一种自适应方式,但严格地说是水印的自适应生成方式。自适应嵌入方法包括:乘性自适应、加性自适应、块自适应和自嵌入等。

3. 水印检测技术

数字水印的检测算法和提取算法是数字水印系统的关键部分之一。水印检测是水印嵌入的逆过程。所谓水印检测,是指根据检测密钥通过一定的算法判断可疑作品中是否含水印。所谓水印提取,是指根据提取密钥通过一定的算法(往往是嵌入算法的逆过程)提取出可疑作品中的每个印记,其长度等于原始水印序列的长度。水印提取的输出

结果是水印，而水印检测的结果是判断水印是否存在的判决结果。水印提取/检测算法的流程如图 3.15 所示。如果水印检测或提取过程中需要用到原始载体，则称此过程为明检测或明提取；如果水印检测或提取过程中不需要用到原始载体，则称此过程为盲检测或盲提取。一方面，水印提取过程往往与水印嵌入算法密切相关；另一方面，水印提取之前往往先进行水印检测（也有文献采用先提取水印后用相关方法判断水印的有无）；此外，有些文献甚至不区分这两个概念，而统称为水印检测；也有的文献采用水印恢复或水印解码（由提取的印记恢复所嵌入的消息）的概念。在水印的提取过程中，宿主图像是可选的，它取决于水印的嵌入算法。当水印算法为盲检测时，检测过程不需要宿主图像。反之，需要宿主图像。盲检测水印算法使用方便，实用性更强。下面侧重讨论水印检测问题。检测算法的设计依赖于嵌入器的嵌入规则。一个水印检测算法是否有效，取决于人们所选取的检测算法的模型与实际是否接近，检测算法的模型包括一系列的假设，主要有以下三个方面：① 对待检测信号的统计特性的假设，在水印系统中对应于水印信号的统计特性。② 噪声的统计特性的假设，在水印系统中对应于载体信号和攻击引入的噪声的统计特性。③ 噪声与信号的叠加方式，即水印的嵌入方式。

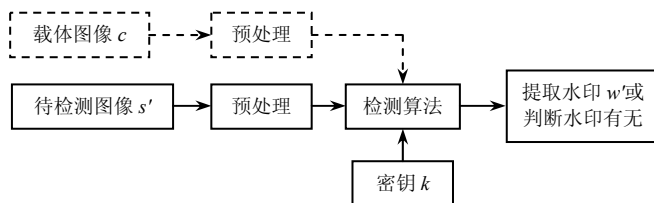


图 3.15 水印提取/检测流程

当嵌入到载体作品中的水印是一个伪随机序列时，水印检测器只需要做出有无水印的判断。在另外一些应用中，需要将视觉可辨的图案嵌入到载体作品中，嵌入的信息首先转换为符号序列，然后水印嵌入器将每个符号嵌入到载体作品的一段中。在这种情况下，检测器首先需要检测出每段作品中嵌入的符号，然后恢复出整个水印，最后针对提取后的水印图像作出有无水印的判断。第一种情况只需要水印检测过程。第二种情况下提取每个符号的过程类似于第一种情况的检测过程，但它们不完全一样，因为在选取检测阈值时需要考虑不同的因素。另外，第二种情况需要对最后提取的水印信息进行处理并作出最终判断。第一种情况可看成第二种情况的特例，即嵌入了 1 比特水印信息。

（1）相关检测

相关检测的主要思想是计算接收到的载体作品 s' 与水印信号 w 之间的相似性，通过相似性度量是否超过给定阈值来判断载体作品 s' 中是否已经嵌入水印 w 。由信号检测理论可知，当信号中叠加的噪声是加性高斯白噪声时，如果把最大化检测器输出端的信噪比作为优化目标，此时相关检测器是最优检测器。盲检测算法中，水印检测器不需要原始载体作品参与检测过程。相关检测器计算作品 s' 和水印信号 w 之间的线性相关值

$$z = \frac{1}{N} \sum_{i=1}^N s'(i) \cdot w(i) \quad (3.21)$$

将检测结果与预先设定的阈值 z_r 相比较，根据检验统计量 z 是否超过检测阈值 z_r 来确定载体作品中是否嵌入了水印。当 $z \geq z_r$ ，判定 s' 中有水印，当 $z < z_r$ ，判定 s' 中没有水印。相关检测器的结构如图 3.16 所示。

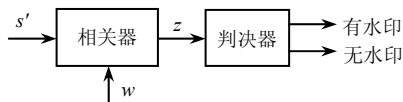


图 3.16 相关检测器的结构

(2) 假设检验方法

使用假设检验方法检测水印主要考虑到以下原因。① 人们可以获得载体作品的完整的统计知识——概率密度函数。在相关检测中假定它服从高斯分布，从而只需要知道载体作品的一阶、二阶矩。人们应该充分利用载体作品的统计分布知识，导出最优统计量和检测阈值，降低错误率。② 在水印嵌入器中，为了提高鲁棒性同时保持水印的不可感知性，人们都采用了视觉或者听觉隐蔽模型。对于添加到作品中的水印信号，嵌入器根据载体信号的特性调整嵌入强度。这种情况下，水印信号的时域、频域特性与载体作品紧密相关，在检测端，不能将载体作品看成独立于水印的加性噪声。

在上述情况下，可以采用假设检验方法检测水印。设要检测的作品 s' 属于以下两种情况之一

$$\begin{aligned} H_0: s' &= c \\ H_1: s' &= c + w \end{aligned} \quad (3.22)$$

即假设 H_0 假定接收到的作品 s' 中没有嵌入水印，称为零假设。假设 H_1 假定接收到的作品 s' 中嵌有水印 w ，称为备择假设。满足假设 H_0 的作品全体组成集合 R_0 ，满足假设 H_1 的作品全体组成集合 R_1 。假设检验方法，顾名思义，就是利用接收到的载体作品对假设进行检验，从而判断假设是否成立。既然检测器需要判定输入的作品 s' 属于两种假设中的哪一个，那么就需要有一个判别准则，水印检测器的判别准则主要有两个：最大后验概率准则和聂曼-皮尔逊准则，读者可以查阅相关文献进行深入了解，在此不再赘述。

3.5.4 数字图像水印算法的评价

透明性、鲁棒性和安全性是评价数字图像水印算法的三大重要指标，分别描述如下。

1. 透明性（不可见性）

透明性指数字水印的嵌入不影响图像的正常使用，不会引起视觉感官上的质量下降，人们无法得知图像中水印的存在。数字水印应是不可知觉的，而且应不影响被保护数据的正常使用。对水印可见性的测量可采用定量方法（Quantitative Metric）或主观测试方法（Subjective Test）。前者主要包括两类方法：一种是基于像素的测量方法；另一种采用基于人类视觉系统的度量方法。而在主观测试方法中介绍一种基于观测者打分的质量等级评判方法。这里只介绍主观打分和基于像素的客观方法。

(1) 主观评价

对视觉质量的评价可以采用主观打分的方法进行。主观评价是检验水印透明性的首要标准。人是图像信息的最终接收者，因此从视觉上要求嵌入水印的图像与原图差别很小，人无法从含水印图像中看出水印。当进行主观测试时，必须遵循一个测试协议，此协议描述了测试和评价的整个过程。这种测试通常分成两步：第一步是将有失真的数据集按由好到坏的次序分成几个等级；第二步，测试者被要求给每个数据集打分和根据降质情况描述可见性。这种打分也可基于 ITU-R Rec.500 质量等级评判（表 3.2 所列出的评分标准）。这些实验从本质上来说都是属于统计性的，不同的观察者的表现会有所不同：

一位观察者也许会说它在两幅作品之间看到了一点差异，而另一个观察者则可能没有发现。有些时候这些差异带有随机性，比如专业摄影师和研究人员对水印图像的主观测试结果差异很大。而对于任何一个人来说，这些灵敏度也会随着时间的变化而变化。研究的结果也许只是针对样本所选取的特定群体（例如，18~25 岁的年轻人）。主观测试对最终的图像质量评价和测试是十分有用的。但是，在研究和开发中，用处却并不大。

表 3.2 ITU-R Rec.500 中定义的品质和作品失真等级

五个等级的标准			
品 质		作 品 失 真	
5	优秀	5	不可感知
4	良好	4	可感知，但不让人厌烦
3	一般	3	轻微的让人厌烦
2	差	2	让人厌烦
1	很差	1	非常让人厌烦

（2）基于像素的客观评价

目前常用的图像质量测试方法是基于图像像素亮度值的。基于像素的失真度量方法属于定量度量法（Quantitative Distortion Metric），用它得到的结果不依赖于主观评价，它允许在不同的方法之间进行公平的比较。大部分视觉信息处理中的失真度量或质量度量方法都属于差分度量法（Difference Distortion Metric）。例如，要测试两幅图像的差别，一般用两幅图像像素亮度值之间的峰值信噪比（PSNR）或相关函数来表征它们之间的差别。表 3.3 给出了常用的各种基于图像像素的测试公式。表中， c_{ij} 表示大小为 $A \times B$ 原始图像的位置 (i, j) 的像素亮度值， s_{ij} 表示对应的含水印图像像素亮度值。从表面上看，这种测试方法似乎可给出定量测试值。但是，由于它不是基于人类视觉模型的测试方法，有一定的局限性。尤其对彩色图像的质量测试，它不能表示出彩色色度的变化和失真程度，往往看上去不错的图像，可能其 PSNR 值却很小，而看上去不太好的图像，其 PSNR 值却很大。因此，基于图像像素亮度值差异的图像质量测试方法是有欠缺的，必须进行改进。近几年，越来越多的研究集中于适合于人类视觉系统的具有自适应性能的失真度量方法，最近的一些数字水印系统测试基准软件已经开始使用这种定量度量方法。

2. 鲁棒性

鲁棒性指在经过常规信号处理操作后能够检测出水印的能力。在某些情况下，鲁棒性毫无用处甚至被极力避免，如脆弱水印要求对图像做任何信号处理操作都会将水印破坏掉。各种不同的水印系统最重要的性能之一就是鲁棒性，而一般来说鲁棒性与不可见性之间存在着矛盾。因此，一般必须在给定水印图像视觉可见性的情况下来研究水印系统的鲁棒性。同时，鲁棒性还与嵌入数据量、水印嵌入强度等因素有关。通常水印的嵌入量越大，鲁棒性越好，但是透明性就越差。一般来说，水印鲁棒性与应用目的类型无关，主要依赖于下面几个方面。

（1）嵌入信息的数量

这是一个重要的参数，因为它直接影响水印的鲁棒性。对同一种水印方法而言，要嵌入的信息越多，水印的鲁棒性越差。被嵌入的信息依赖于各种应用场合。为解决所有权问题，可隐藏类似于 ISBN（约 10 个数）或 ISRC（约 12 个字母、数字、字符）这样的数字，在此数之前可能还需加上版权年份、著作允许权及允许级别等，这就意味着在一

幅图像中需嵌入约 70 位的秘密信息，这还不包括用于纠错编码所附加的比特数。

(2) 水印嵌入强度

水印嵌入强度（对应于水印的鲁棒性）和水印可见性之间存在着一个折中。增加鲁棒性就要增加水印嵌入强度，相应地会增加水印的可见性。

表 3.3 常用的基于图像像素的图像质量测试公式

测试方法	公 式
最大差值	$MD = \max_{1 \leq i \leq A, 1 \leq j \leq B} s_{ij} - c_{ij} $
平均绝对差值	$AD = \frac{1}{A \times B} \sum_{i=1}^A \sum_{j=1}^B s_{ij} - c_{ij} $
归一化平均绝对差值	$NAD = \frac{\sum_{i=1}^A \sum_{j=1}^B s_{ij} - c_{ij} }{\sum_{i=1}^A \sum_{j=1}^B c_{ij} }$
均方差值	$MSE = \frac{1}{A \times B} \sum_{i=1}^A \sum_{j=1}^B s_{ij} - c_{ij} ^2$
归一化均方差值	$NMSE = \frac{\sum_{i=1}^A \sum_{j=1}^B s_{ij} - c_{ij} ^2}{\sum_{i=1}^A \sum_{j=1}^B c_{ij} ^2}$
L_p 范数	$L_p = \left(\frac{1}{A \times B} \sum_{i=1}^A \sum_{j=1}^B s_{ij} - c_{ij} ^p \right)^{1/p}$
Laplacian 均方差值	$LMSE = \frac{\sum_{i=1}^A \sum_{j=1}^B \nabla^2 s_{ij} - \nabla^2 c_{ij} ^2}{\sum_{i=1}^A \sum_{j=1}^B \nabla^2 c_{ij} ^2}$ $\nabla^2 c_{ij} = c_{(i+1)j} + c_{(i-1)j} + c_{i(j+1)} + c_{i(j-1)} - 4c_{ij}$
信噪比	$SNR = 20 \lg \left(\frac{\sum_{i=1}^A \sum_{j=1}^B c_{ij} ^2}{\sum_{i=1}^A \sum_{j=1}^B s_{ij} - c_{ij} ^2} \right)$
峰值信噪比	$SNR = 20 \lg \left(A \times B \max_{1 \leq i \leq A, 1 \leq j \leq B} c_{ij} ^2 / \sum_{i=1}^A \sum_{j=1}^B s_{ij} - c_{ij} ^2 \right)$
图像保真度	$IF = 1 - \frac{\sum_{i=1}^A \sum_{j=1}^B s_{ij} - c_{ij} ^2}{\sum_{i=1}^A \sum_{j=1}^B c_{ij} ^2}$
归一化互相关系数	$NC = \frac{\sum_{i=1}^A \sum_{j=1}^B (s_{ij} \times c_{ij})}{\sum_{i=1}^A \sum_{j=1}^B c_{ij} ^2}$
相关品质	$CQ = \frac{\sum_{i=1}^A \sum_{j=1}^B s_{ij} \times c_{ij} }{\sum_{i=1}^A \sum_{j=1}^B c_{ij} }$
构造内容比	$SC = \frac{\sum_{i=1}^A \sum_{j=1}^B c_{ij} ^2}{\sum_{i=1}^A \sum_{j=1}^B s_{ij} ^2}$

(3) 图像的尺寸和特性

图像的尺寸对嵌入水印的鲁棒性有直接影响。尽管非常小的含水印图片没有多少商业价值，但一个水印软件程序应该能够从此图片中恢复出水印。对于用于打印的图像，常常想要高解析度的。但是，同时也希望这些图像被采样并被放到万维网上以后能够得到保护。除图像尺寸之外，图像的特性也对水印的鲁棒性产生重要影响。

(4) 水印攻击类型

在对水印系统进行性能评估的过程中，需要对水印系统进行一些攻击，以测试其性能。这些攻击是一个水印系统在实际使用过程中可能会遭受到的，此处“攻击”的含义包括有意的攻击和无意的攻击。有意的攻击是指为了去除水印而采取的各种处理方法，此种攻击往往是恶意的，具体见第 8 章；无意的攻击是指加有水印的图像在使用过程中不可避免受到的诸如有损压缩、噪声影响等处理。下面将一一列出各类攻击。

① JPEG 压缩攻击。JPEG 是广泛用于图像压缩的压缩算法，通常经过水印系统处理的图像必须能经受某种程度的有损压缩，并能提取出压缩后含水印图像中的水印。

② 几何失真攻击。几何失真包括下述各种几何操作。

- 水平翻转。许多图像可以被翻转而不丢失数据，尽管对翻转的抵御很容易实现，但却很少有系统能够真正逃脱这种攻击。
- 旋转。一般进行小角度旋转（通常混有剪切）并不会改变图像的商业价值，但却能使水印无法检测出来。
- 剪切。对图像进行剪切可破坏水印，这在某些情况下很有用处。有时候，盗版者仅对有版权保护的原始图像的中央部分感兴趣。此外，越来越多的 Web 站点使用图像分割方法，这造成了 Mosaic 攻击方法的产生。
- 尺度变换。在扫描打印图像时或在将高分辨率数字图像用于 Web 发布时，常会带来尺度的变换。尺度变换可分为两类：一致尺度变换和非一致尺度变换。一致尺度变换是指在水平方向和垂直方向进行相同的尺寸变换，而非一致尺度变换指在水平和垂直方向使用不同的尺度因子（采用不同的比率）。通常水印方法只能抵御一致尺度变换。
- 行列删除。此方法对于攻击在空域上直接运用扩频技术嵌入水印的算法十分有效。
- 广义几何失真。广义几何失真是非一致尺度变换、旋转和剪切的综合。
- 随机几何失真。这是 StirMark 中使用的方法，该方法模拟了一幅图像经高质量打印机打印后再扫描进计算机所带来的噪声和失真。具体来说，首先对图像进行微小随机量的几何失真：轻微的拉伸、扭曲、平移和旋转，然后采用双线性插值或奈奎斯特插值方法进行重采样，在此过程中，对所有像素点通过一个传递函数引入微小的分布误差。
- 和 JPEG 结合进行几何失真。单独使用旋转、尺度变换并不够，它应和 JPEG 压缩结合起来，对水印技术进行测试。由于大多数情况下会先对图像进行几何变换，然后再用压缩格式保存图像，这就使得测试水印系统对由压缩带来的几何失真的鲁棒性很有意义。选择一个合适的 JPEG 压缩质量因子是一个重要的问题，因为随着质量因子的减小，降质会迅速出现，实验表明不低于 70% 的质量因子是合适的。

③ 增强处理攻击

- 低通滤波。包括线性和非线性滤波器。经常使用的滤波器有中值滤波、高斯滤波和标准均值滤波。
- 锐化。锐化处理属于标准图像处理，这种处理可用作对水印系统的有效攻击，因为它们在检测由数字水印软件带来的高频噪声方面十分有效。更加细微的攻击是建立在拉普拉斯算子基础上的。最简单方法就是对含水印图像 s 采用如下算子进行处理

$$s' = s - \alpha \nabla^2 (\nabla^2 s - s) \quad (3.23)$$

其中， α 是攻击强度。

- 直方图修正。包括直方图拉伸或均匀化，直方图均匀化常用来对照明条件较差的图像进行补偿处理。
- Gamma 校正。这是一种经常使用的方法，常用来增强图像或调整图像使其适合于显示，例如在扫描后经常进行 Gamma 校正。

- 颜色量化。这通常用于将真彩图像转换成 GIF 格式图像时。颜色量化通常需要进行抖动处理, 这种处理扩散了由量化带来的误差。
- 复原。在图像处理中, 这类技术常用来减小某些特定的降质过程带来的图像降质。采用此种方法处理水印图像不需要知道水印系统噪声的先验知识。

④ 附加噪声攻击。在图像传送和处理过程中, 存在着大量的加性噪声和非相关的乘性噪声。许多水印系统能够抵御这类噪声, 但存在一个可接受的干扰噪声的最高限度。

⑤ 打印扫描攻击。这个过程将引入几何失真和类似噪声的失真。

⑥ 统计平均和共谋攻击。若能获得同一图像的多个复制, 但每幅图像带有不同的水印, 则可通过对这些图像进行平均或取出所有图像的一部分进行重新组合来去除水印。

⑦ 嵌入多重水印攻击。就是在已经加有水印的图像中再嵌入一个水印。

⑧ Oracle 攻击。有时水印解码器是公开给所有人使用的, 此时攻击者可不断地对含水印图像作微小修改直到水印解码器不能检测出水印为止, 以此来删除水印。

鲁棒性评价由提取水印与原始水印间的差异(相似度)来衡量, 当含水印图像没有受到攻击时, 相似度为 1 或者接近 1, 表明水印几乎没有失真。而当含水印图像受到攻击时, 提取的水印与原水印的相似程度则用归一化相关系数(Normalized Correlation, NC)来评价。NC 值越接近于 1, 则表明提取的水印就越接近于原始水印, 水印的鲁棒性就越好。若嵌入的水印信号 w 是一幅 $A \times B$ 大小的图像, 则 NC 的计算公式为

$$NC = \frac{\sum_{i=1}^A \sum_{j=1}^B w_{ij} w'_{ij}}{\sum_{i=1}^A \sum_{j=1}^B w_{ij}^2} \quad (3.24)$$

其中, w 表示原始水印, w' 表示提取出的水印。

3. 安全性

安全性指水印能够抵抗各种破坏水印功能的行为的能力, 即未授权者不能去除、嵌入和检测水印。同时, 水印信息应该很难被他人所复制和伪造。现代密码体制要求满足 Kerckhoffs 准则, 即算法的安全性要取决于密钥的安全性, 而不是建立在整个算法保密的基础上。人们希望水印算法也能满足 Kerckhoffs 准则。如果密钥未知, 即使水印算法已知。攻击者也不可能检测出图像中是否有水印。

3.5.5 典型鲁棒图像水印算法

鲁棒水印的特点是水印信息难以被去除, 主要用于版权保护。1994 年, Van Schyndel 等首先提出了空域的最低有效位(LSB)算法^[34]。LSB 算法的优点是算法简单、嵌入容量大、不可见性好和提取信息时不需要宿主图像等优点。但是, 该算法实质上相当于在图像中添加一些高频噪声, 对图像的几何变形和信号处理如滤波、压缩、加噪声等抵抗能力差, 鲁棒性不好, 不能应用于实际。1996 年, Bender 等人提出了著名的 Patchwork 算法^[16]。这是一种基于统计的算法, 即在一个载体图像中嵌入具有特定统计特性的水印, 其嵌入方法是任意选择一对像素点, 在增加某像素亮度的同时, 降低另一像素的亮度值。该算法的隐藏性较好, 并且对有损的 JPEG 和滤波、压缩和扭转等操作具有抵抗能力, 但仅适用于具有大量任意纹理区域的图像。LSB 算法和 Patchwork 算法属于空间域水印的经典算法。通常, 空域水印算法通过修改载体图像部分像素的灰度值很

容易嵌入水印，但嵌入的水印信息很容易通过分析而检测出来。变换域水印算法主要通过修改图像的变换域（DFT、DCT 或 DWT）系数来嵌入水印。结合人类视觉系统（HVS）特点，为保证水印算法的鲁棒性，水印信息应该嵌入到变换域的低频和中频系数中。但是在频域中，通常不能嵌入太多的数据，否则会导致载体图像的严重失真。也就是说，水印的尺寸必须小于载体图像，一般而言，它是载体图像的 1/16。

为了嵌入更多的水印，降低含水印图像的失真，文献[35]提出一种将时域和空域结合起来的技术。根据用户的偏好和图像数据重要性，将水印图像分成时域和空域两部分。该水印算法具有如下优点：可在载体图像中嵌入更多的水印，使鲁棒性增强。把水印分裂成两部分，使得安全程度加倍。水印分裂方法还可设计得更加复杂以至于很难被破解。此外，为增强鲁棒性，水印的随机排序也被用来抵抗信号处理（如图像剪切）的攻击。为描述方便，下面称该算法为组合水印算法。

组合水印算法的主要思想是将水印图像分裂成两部分，一部分嵌入到载体图像的空域中，另一部分嵌入到载体图像的频域中。设 c 表示大小为 $A \times A$ 的载体灰度图像， w 表示大小为 $B \times B$ 的二值水印， $w^{(1)}$ 和 $w^{(2)}$ 是 w 分裂出来的两个水印， $s^{(1)}$ 是在空域把 $w^{(1)}$ 嵌入到 c 后得到的图像， $S^{(1)}$ 是 $s^{(1)}$ 经过 DCT 变换后得到的结果。 $S^{(2)}$ 是在频域把 $w^{(2)}$ 嵌入到 $S^{(1)}$ 后的得到的结果。 $S^{(2)}$ 经过 IDCT 输出最终的含水印图像 s 。 \oplus 表示替代载体图像 LSB 的操作。组合水印算法的流程图如图 3.17 所示。

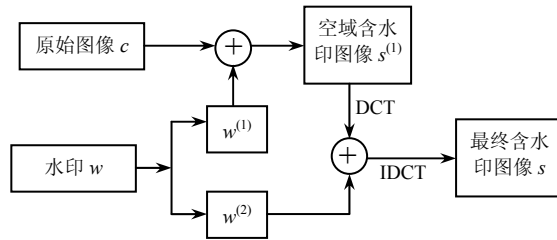


图 3.17 空域和频域组合水印算法流程图

组合水印算法可描述如下。

(1) 把水印 w 分裂成空域和频域两部分（以便分两步嵌入到输入图像中），即将二值水印 $w = \{w_{ij} | w_{ij} \in \{0,1\}, 0 \leq i, j < B\}$ 分成 $w^{(1)} = \{w_{ij}^{(1)} | w_{ij}^{(1)} \in \{0,1\}, 0 \leq i, j < B_1\}$ 和 $w^{(2)} = \{w_{ij}^{(2)} | w_{ij}^{(2)} \in \{0,1\}, 0 \leq i, j < B_2\}$ 两部分，其中 $B^2 = B_1^2 + B_2^2$ 。它的分裂标准依赖于用户和水印的应用场合。原则上，最重要的数据存在于图像的中间，最好嵌入到频域中。因此，一个简单的分裂方法就是选择水印图像的中心窗口数据嵌入到频域中。

(2) 空域水印嵌入。将 $w^{(1)}$ 嵌入到 c 的空间域中，获得空域含水印图像如下

$$s^{(1)} = \{s_{ij}^{(1)} | s_{ij}^{(1)} = c_{ij} \oplus w_{ij}^{(1)}, 0 \leq i, j < A\} \quad (3.25)$$

其中， $c_{ij}, s_{ij}^{(1)} \in \{0,1,2,\dots,2^L-1\}$ ， L 是像素的灰度级。式中 \oplus 表示用水印图像的像素去替换载体图像像素的最低位。

(3) 对 $s^{(1)}$ 进行 DCT 变换得到 $S^{(1)}$ 。

(4) 将 $w^{(2)}$ 嵌入 $S^{(1)}$ 中，获得变换域含水印图像 $S^{(2)}$ 如下：

$$S^{(2)} = \{S_{ij}^{(2)} | S_{ij}^{(2)} = S_{ij}^{(1)} \oplus w_{ij}^{(2)}, 0 \leq i, j < A\} \quad (3.26)$$

设 $s^{(1),u}$ ， $w^{(2),v}$ 分别是 $s^{(1)}$ 和 $w^{(2)}$ 的子块； $S^{(1),u}$ 是 $s^{(1),u}$ 的 DCT 变换； $S^{(2),u}$ 是 $S^{(1),u}$ 嵌入 $w^{(2),v}$ 的结果，则式 (3.26) 所描述的算法具体过程如下：① 将图像 $s^{(1)}$ 分解成 8×8 的子

块 $s^{(1),u} = \{s_{ij}^{(1),u}, 0 \leq i, j < 8\}$, u 是 8×8 子块的序号。② 将水印图像 $w^{(2)}$ 分解成 2×2 的子块 $w^{(2),v} = \{w_{ij}^{(2),v}, 0 \leq i, j < 2\}$, v 是 2×2 子块的序号。③ 对 $s^{(1),u}$ 进行 DCT 变换得到 $S^{(1),u}$ 。④ 利用 LSB 替换方法将 $w^{(2),v}$ 嵌入到 $S^{(1),u}$ 中的四个中频系数中, 得到 $S^{(2),u}$ 。在频域中嵌入水印的一个要求是图像 $s^{(1)}$ 的 8×8 的子块数目必须比水印图像 $w^{(2)}$ 的 2×2 子块总数多。

(5) 对变换域含水印图像 $S^{(2)}$ 进行 IDCT 变换得到最终的含水印图像 s 。

为了验证上述方法的有效性, 文献[39]给出了一些仿真试验。图 3.18 说明了水印对于安全性的重要性。比如, 若要禁止人们看到图像中的作者名, 则可把它截下来嵌入到频域中, 其余嵌入到空域中, 通过这种方法, 不但扩大了水印容量, 也增加了人们所关心信息的安全性。

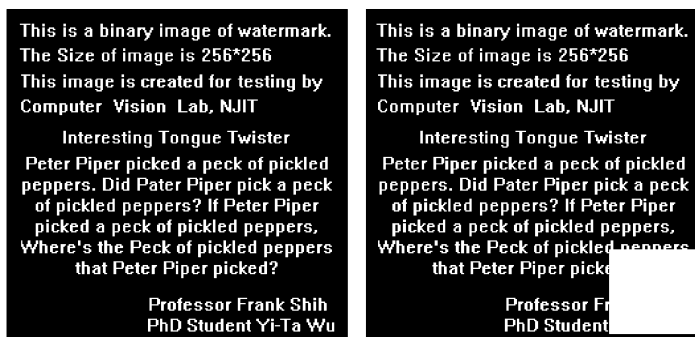


图 3.18 从一幅 256×256 的图中截取大小为 64×64 的图像

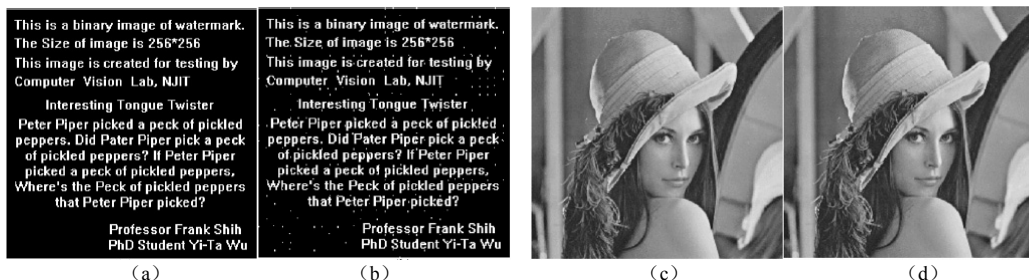
图 3.19 表示大小为 256×256 的原始 Lena 图像。图 3.20 给出用传统 LSB 替换技术在载体图像频域中嵌入 64×64 水印的结果, 其中图 (a) 为原始水印, 图 (b) 是嵌入该水印得到的 Lena 图像 ($\text{PSNR}=64.57\text{dB}$), 图 (c) 为从 (b) 中提取的水印 ($\text{NC}=1$)。图 3.21 给出用组合算法嵌入大小为 128×128 的水印的实验结果。图 3.21 (a) 为大小为 128×128 的原始水印。该水印分解为图 (b) 和图 (c) 两部分, 图 (e) 和 (f) 分别是由组合算法得到空域含水印图像和最终的含水印图像, PSNR 分别为 56.58dB 和 55.93dB 。图 3.21 (d) 是从图像 3.21 (f) 中提取的水印, $\text{NC}=0.9813$ 。图 3.22 (a) 给出更大的水印, 大小为 256×256 , 该水印按图 3.18 分解为两部分。图 3.22 (c) 和 (d) 分别是由组合算法得到空域含水印图像和最终的含水印图像, PSNR 分别为 51.14dB 和 50.98dB 。图 3.22 (b) 是从图像 (d) 中提取的水印, $\text{NC}=0.9644$ 。



图 3.19 原始 Lena 图像



图 3.20 用 LSB 替换在载体图像频域中嵌入 64×64 的水印
(a) 原始水印; (b) 含水印图像; (c) 抽取的水印

图 3.21 用组合法嵌入大小为 128×128 的水印的实验结果图 3.22 用组合法嵌入大小为 256×256 的水印的实验结果

3.5.6 典型脆弱图像水印算法

所谓脆弱数字水印技术就是在保证一定视觉质量前提下，将数字水印嵌入到多媒体数据中，当多媒体内容受到怀疑时，提取该水印来鉴别多媒体内容的真伪，并指出篡改位置，甚至攻击类型等。脆弱水印的特点是水印信息可以随着嵌入水印后图像的破坏而被破坏，主要用于图像完整性验证。脆弱性水印是指水印能够检测出对图像做出的任何改动，甚至 1 比特。由于图像数据在信道传输过程中，容易受到各种噪声的污染，且传输时往往进行图像压缩。因此，脆弱水印算法的实用性不强。

早期的脆弱水印技术大多由 LSB 算法演变而来，且算法简单，易于硬件实现。改进后的脆弱水印算法大多在变换域内实现，并有了新的特性，算法本身不但可以实现对篡改位置的定位，甚至还可以用于篡改部位的复原。下面介绍文献[36]提出的一种基于 DCT 系数量化的用于图像认证的安全脆弱水印技术，其基本步骤可描述如下。

(1) 去除图像像素间的相关性。这样做有两个主要目的。第一，有助于增加安全性，即防止嵌入的水印被非法提取。第二，对于 DCT 变换来说，这种做法可引入大量中高频成分，从而增加了可用于嵌入水印的变换系数个数。设 c 为待认证的原始图像， $c' = c - \mu$ 为去除平均值后的同一幅图像，其中 μ 代表均值图像（各像素的亮度均为图像的平均亮度）。定义 F 为取决于密钥 $k \in K$ 的混乱算子，其中 K 为密钥空间，则像素位置混乱后的图像就是去除相关性后的图像 c'' （去除图像的相关性等同于白化处理，也就是使图像看起来像白噪声图像一样），表示如下

$$c'' = F(c', k) = F(c - \mu, k) \quad (3.27)$$

(2) 分块 DCT 变换和量化。把图像 c'' 分成一系列大小为 $A \times A$ 的小块 $\{c_u''\}$ ，对每块 c_u'' 分别作 DCT 变换得到 DCT 系数块 C_u'' ，这里下标 u 表示块的序号。接着， C_u'' 的每个 DCT 系数 $C_{u,ij}$ 的幅度量化成最接近的整数倍 $2^n A$ ，其中 n 代表每个变换系数要嵌入的比特数，下标 i 和 j 表示系数在块内的位置。选择较小的量化阶矩 A 使含水印图像与原始

图像保持视觉上的相似性。

(3) 嵌入水印。对于每个量化后的 DCT 块 $C_u''^Q$ ，相应的变换域含水印图像块可通过对量化后的每个系数 $C_{u,ij}''^Q$ 中直接加入水印信息 $w_{u,ij}$ 获得，即

$$S_{u,ij} = C_{u,ij}''^Q + w_{u,ij} \quad (3.28)$$

其中， $0 \leq i, j < A$ ， $0 \leq |w_{u,ij}| \leq 2^n \Delta$ 。若每个变换系数量化后用来嵌入 2 比特水印，即 $n=2$ ，则水印位可映射到 4 个互不重叠的量化电平，每一个电平值为

$$Q_k = q \cdot \Delta + \frac{\Delta}{2}, \quad q \in \{0, 1, 2, 3\} \quad (3.29)$$

于是，每一个含水印系数都可表示为 $2^n \Delta$ 乘以整数 l 再加上嵌入的水印信息，即

$$S_{u,ij} = C_{u,ij}''^Q + w_{u,ij} = l \cdot 2^n \Delta + w_{u,ij} \quad (3.30)$$

其中 $w_{u,ij} \in \{Q_0, Q_1, Q_2, Q_3\}$ 。

(4) 得到含水印图像。嵌入过程完成之后对各块 S_u 进行 DCT 逆变换，拼接成图像后加入原来的均值图像 μ ，然后用同样的密钥 k 进行逆混乱即可得到含水印的图像 s 。

水印提取过程比较简单。首先，把接收到的图像 s' 的平均亮度去除。然后，对得到的图像用密钥 k 进行混乱以去除像素间的相关性。接着，对去相关图像进行分块 DCT 变换，得到 S_u' ，这里下标 u 表示块的序号。然后，对各块 S_u' 的系数值经过模运算后用再用阶距 Δ 量化即可提取水印。提取过程的表达式如下

$$w'_{u,ij} = Q[\text{mod}(|S'_{u,ij}|, 2^n \Delta)] \Delta \quad (3.31)$$

式中，下标代表被 Δ 量化。量化结果为 $2^n=4$ 个不同电平，每一个电平对应嵌入到特定 DCT 系数中的实际位值，即式 (3.29) 中的 q 值对应的 2 比特二进制数。

3.6 数字音频水印技术

与数字图像水印技术相比，数字音频水印技术有自己的特性：① 音频信号在每个时间间隔内采样的点数要少得多，这就意味着音频信号中可嵌入的信息量要比可视媒体少得多；② 人耳的听觉系统 (HAS) 要比人眼视觉系统 (HVS) 灵敏得多，因此听觉上的不可感知性实现起来要比视觉上困难得多；③ 为了抵抗剪切攻击，嵌入的水印应该保持同步；④ 由于音频信号一般都比较长，所以提取是不能需要原始音频的；⑤ 音频信号有特殊的攻击，如回声、时间缩放等。因此与数字图像水印相比，数字音频水印具有更大的挑战性。下面首先介绍音频水印系统的基本要求，然后介绍音频水印系统的基本模型，接着介绍含水印音频的质量评价问题，然后介绍数字音频水印系统的鲁棒性评测问题，最后分时间域、变换域和压缩域分别介绍一些典型的数字音频水印算法。

3.6.1 音频水印系统的基本要求

对音频水印系统的要求与其他水印系统类似，主要包括鲁棒性、透明性、安全性以及对载荷 (容量)、计算复杂性、检测器错误率的要求。

1. 透明性

透明性要求在高品质音乐制品版权保护中是必需的，它要求含水印音乐制品与原始

音乐制品之间不存在感知上的差异，并且二者经过相同的信号处理操作后也不应该存在感知差异。在另外一些应用中，例如隐秘语音通信，这种要求会有些放松。

2. 鲁棒性

含水印音乐制品在分发和传递过程中会不可避免地受到各种信号处理操作。鲁棒性的含义是：除非信号处理操作严重降低了载体音频作品的品质，否则经过信号处理操作后，水印检测器应该仍能检测出载体作品中是否含水印。常见的信号处理操作包括：添加噪声、线性和非线性滤波、有损压缩、时间轴缩放、DA/AD 转换等。已经有若干关于音频水印鲁棒性的测试标准，这部分内容将在音频水印攻击中详细讨论。

3. 安全性

同其他媒体水印技术一样，音频水印系统必须考虑其应用环境中可能受到的攻击。当水印技术用于版权保护时，一个重要的问题是如何确定版权所有者，特别是当音频作品中含有多个水印，各方都声称具有作品版权时，如何解决这种“死锁”问题等。

4. 载荷

不同应用环境对载荷的要求不同。在版权保护应用中，往往只需要把序列号和作者的标识码重复地嵌入到载体作品中，这种情况下对载荷的要求不高。在使用水印技术进行隐秘通信应用中，对信息载荷的要求一般很高。

5. 计算复杂性

计算复杂性的要求也是随着应用环境的不同而不同，一般而言，要求检测器的复杂性要低于嵌入器的复杂性。在诸如广告监控等应用中，要求算法应具有较低的复杂性，从而有利于实时实现。使用心理感知模型往往会增加复杂性，这是由于不仅在水印嵌入阶段而且在水印检测阶段也需要使用心理声学计算模型。

6. 错误率

在版权保护应用中，为了增加证据的说服力，一般要求水印检测器具有较低的虚警率，特别是在将检测结果作为法律纠纷中法庭的证据时，这种情况下往往使用错误率作为设计指标来决定其他参数的选取，例如使用虚警率决定相关检测的检测阈值的选取。

由上述的讨论可知，对音频水印系统的要求随着应用环境的不同而不同，各个要求之间可以互相折中，所以水印设计者不应该对水印系统提出不切实际的要求，而应根据实际应用环境寻找合理的要求。

3.6.2 音频水印系统的基本模型

水印模型是水印算法基础，音频水印中常用的算法模型与图像水印等类似，图 3.23 给出了针对音频水印的算法模型。其中，水印编码环节负责秘密信息加密、纠错编码。心理声学模型提供掩蔽阈值信息来确定水印的最大可能嵌入强度。水印提取环节与水印嵌入环节相对应。水印检测环节负责判断水印的存在性，但是不提供水印的内容。水印解码环节负责提取秘密信息。这个模型是一个基本模型，具体水印算法可能只包含其中的几个环节，例如回声隐藏模型没有明显包含心理声学模型环节，虽然它间接利用了听觉系统的感知特性。另外，若水印本身是一个伪随机序列，则不会包含水印解码环节。

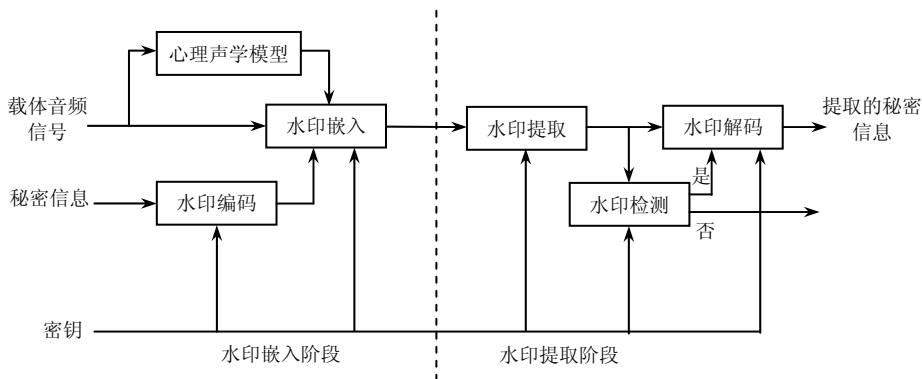


图 3.23 音频水印系统基本模型

3.6.3 含水印音频质量的评价

水印系统的要求之一是水印的不可感知性，所以在水印系统性能测试中需要评价含水印音频信号的感知质量。水印的嵌入会在载体音频信号中引入噪声，该过程与音频编解码器在音频中引入的噪声相似，故对含水印音频质量的评价类似于评价音频编解码器处理后的音频质量。对音频质量的评价包括对音质的评价和对失真的评价，常用的评价方法主要有三种：客观评价方法、主观评价方法和模拟听觉感知特性的客观评价方法。

1. 客观评价方法

客观评价法主要包括时域评价法和频域评价法。时域评价法评价添加水印后的音频与原始音频之间的失真度。最常用的时域失真测度是信噪比（Signal-to-Noise Ratio, SNR）。

$$\text{SNR} = 10 \lg \frac{\sum_{i=0}^{N-1} c(i)^2}{\sum_{i=0}^{N-1} [c(i) - s(i)]^2} \quad (3.32)$$

其中， $c(i)$ 表示原始音频信号， $s(i)$ 表示嵌入了水印后的音频信号， N 为音频采样个数。此信噪比的计算包括了整个音频信号段。该方法的缺点是，高能量段信号的信噪比占主要，不能够反映低能量段信噪比对感知的重要性。由于音频信号，特别是语音信号，是短时平稳信号。所以一个改进的方法是计算短时信噪比，并计算短时信噪比的一个统计值作为失真度，这种方法称为分段信噪比

$$\text{SNR}_{\text{SEG}} = \frac{1}{K} \sum_{k=1}^K \text{SNR}_k \quad (3.33)$$

其中， SNR_k 是第 k 个短时段信噪比。上述的分段信噪比是短时信噪比的统计平均值。研究表明，分段信噪比与听觉系统对音频信号差异的感知是相关的。但是，时域信噪比不能够有效地为人类听觉系统对信号延迟和相位失真的感知特性进行建模，人类听觉系统对信号延迟和相位失真的感知不敏感，而分段信噪比对信号延迟和相位失真十分敏感的。所以不能够使用分段信噪比特性评价回声隐藏和相位编码水印。另外，分段信噪比没有考虑人类听觉系统的感知特性。

上述信噪比的计算是针对整个频带进行的，没有提供关于误差信号在频谱中的分布

情况。为了获得误差信号的频域特性,一般采用滤波器组对原始音频信号和误差信号进行频域分解,然后计算每个子带中误差信号的信噪比。另外一种方法采用计算原始信号和含水印信号功率谱的差异来消除相位对信噪比计算的影响。对于高质量的音频编解码系统而言,其分段信噪比非常依赖于信号的类型并且动态范围很宽,从最低 13dB 到最高 90dB,所以采用客观方法评价音频信号失真并不十分有效。

2. 主观评价方法

由于含水印信号的最终接收者是人,所以主观评价标准是最终的,也是最可靠的标准。常用的主观音质评价方法有 ABX 测试方法、五级平均意见打分 (Mean Opinion Scores, MOS) 和评价音质微损伤的连续打分方法。

(1) ABX 测试

ABX 测试用来测试音频编解码器是否能够提供“透明的”质量,即编解码器引入的噪声对听音者而言是否是“透明的”。使用 ABX 测试评价含水印音频质量的方法如下: A 表示原始音频信号, B 表示含水印音频信号, X 表示可能含有水印也可能不含水印的音频信号。听音者分别听取音频信号 A、B、X,然后判断 X 为 A 或 B。听音者作出正确判断的百分比作为判断含水印音频是否达到“透明”质量的基础。

(2) 五级平均意见打分 (MOS)

MOS 打分常用于评价经过语音编解码器处理后语音信号的质量,它是一种绝对等级打分方法。一般采用五级评分标准,受测试者首先听完被测试声音,然后从五个等级中选取一个等级作为他对所测试声音质量的评价,表 3.4 给出了 MOS 评分方法的标准。MOS 得分为 4~4.5 分被认为是高质量语音,3.5 分左右称为通信质量。

(3) 评价音质微损伤的连续打分方法

为评价高质量音频编解码器处理后音频的质量,国际电信联盟提出一套正式的测试标准 ITU-R Rec.BS.1116,它包括一系列对测试环境和测试过程的规定。这套标准使用“双盲-三激励-参考信号隐藏”的测试方法,激励信号 A 总是原始音频信号(称为参考信号),激励信号 B、C 中包括 A 中的参考信号和处理后的信号(称为受损信号),但是受测试者不知道 B 和 C 中哪一个包含参考信号。测试过程如下。

表 3.4 MOS 评分标准

得 分	质量级别 (MOS)	失真级别 (DMOS)
5	优 (excellent)	不察觉
4	良 (good)	刚有察觉
3	中 (fair)	有察觉,稍微令人讨厌
2	差 (poor)	明显察觉,令人讨厌,但可忍受
1	劣 (bad)	不可忍受

(1) 受测试者听取 A、B、C 三个激励信号。

(2) 受测试者判断 B、C 中哪一个激励信号是隐藏的参考信号,哪一个受损信号。

(3) 受测试者对受损信号采用表 3.5 中的 5 类、41 点评分标准评分,表中的得分单位为 0.1。

(4) 当隐藏的参考信号的平均得分位于受损信号得分的 95%置信区间,且受损信号的平均得分位于参考信号的 95%置信区间时,则认为受损信号达到了“透明”质量。

表 3.5 ITU-R Rec. BS. 1116 音质微损伤评分标准

得 分	失真的绝对级别
5.0	不察觉
4.9~4.0	有察觉, 不令人讨厌
3.9~3.0	稍微令人讨厌
2.9~2.0	令人讨厌
1.9~1.0	十分令人讨厌

主观测试方法一般用于算法标准化过程中, 例如音频编解码器算法标准化过程中需要确定某种新提出的算法是否比已有标准算法性能更优。在水印算法的研究过程中, 一般没有能力进行大规模的主观测试, 基本上采用非正式的、小范围的主观测试。另外, 主观测试方法受测试环境, 受测试者的听力情况等影响, 不具有可重复性。解决这个问题的方法之一是采用客观方法模拟人对音频信号的主观感受。

3.6.4 数字音频水印算法的鲁棒性评测

音频水印技术要完成版权保护、内容认证等功能, 必须具有鲁棒性和安全性, 从而使水印系统能够抵抗普通信号处理操作和恶意攻击。本节主要讨论针对音频水印系统的攻击方法以及评测音频水印系统鲁棒性的标准。

1. 针对音频水印系统的攻击

同对其他水印系统的攻击一样, 可以将针对音频水印系统的攻击分为: 水印去除攻击、去同步攻击、嵌入攻击和检测攻击四类。这里只讨论针对音频水印系统的水印去除攻击和去同步攻击。对水印进行去除攻击主要包括: 采用信号处理操作、针对特定水印算法的去除攻击、共谋攻击、Oracle 攻击等。下面主要讨论针对音频水印系统的信号处理操作攻击、针对特定算法的攻击和去同步攻击。

(1) 针对音频水印的常见信号处理操作攻击

对含水印音乐作品的任何信号处理操作都有可能擦除已经嵌入的水印, 另外, 随着音频信号处理工具在互联网上的普及, 一个没有任何音频信号处理操作知识的人都有能力使用音频信号处理工具实施相应攻击。根据音频信号处理知识, 可以将含水印作品可能受到的信号处理操作分为以下几种。

① 动态范围改变: 主要是采用线性和非线性处理操作改变音频信号的动态范围, 包括改变声压级、限幅和压扩等操作。

② 滤波: 滤波操作改变信号的一部分频谱, 例如低通滤波和高通滤波等。均衡器用来增加或降低某些频谱分量, 也可视为滤波操作。

③ 改变数字音频效果: 这种操作原来主要用于录音过程中, 主要包括添加回声、混响、和声和抖动等, 在常见音频信号处理软件中都提供了上述功能, 所以攻击者也可采用这种后处理方法攻击含水印作品。

④ 格式变化: 包括采样率、量化阶矩的变化以及立体声与单声道信号之间的转换。

⑤ 有损压缩: 有损压缩是音频作品在网络环境下常常受到的一种处理, 因为这样可降低对网络传输带宽的要求。有损压缩与数字水印本身是矛盾的, 有损压缩试图去掉对人类感知而言无关的信号分量, 而音频水印试图将信息隐藏到人类感知不到的信号部分, 故采用基于感知模型的有损压缩算法可去除采用感知整形的水印信号。

⑥ 添加噪声：噪声的添加可以有意的，最严重的情况是攻击者采用心理声学模型对噪声整形后添加到作品中。另外，噪声也可能是其他一些信号处理操作的效果之一。

⑦ 时域拉伸和基音改变。这种操作可在改变基音的同时不改变信号的长度，或者在改变信号长度的同时不改变基音。在使用水印进行广告监控应用中会遇到这种情况，电台会改变信号的长度以适应特定时间间隔的要求。

⑧ 样本置乱：这类攻击包括随机去掉一定数量的样本点和剪切攻击等。

面对如此多的信号处理操作，水印算法设计者在测试算法鲁棒性时需要确定在某个应用环境中可能遇到的操作，这是因为目前为止很难找到一个算法能对上述所有操作具有鲁棒性，一个特定算法应该根据其应用环境来决定应该对何种信号处理操作具有鲁棒性。另一个问题是如何确定可能受到的信号处理操作强度，即以上攻击操作的参数。一个合理假设是：可能受到的信号处理操作应该不会损害音质，故在水印测试中确定信号处理操作的强度时要确定该操作不会使载体作品产生明显的失真。

(2) 针对特定算法的攻击

在水印技术中普遍采用了密码学中的 Kerckhoffs 准则，即水印系统的安全性依赖于密钥的安全性。算法设计者在考虑水印系统的安全性时，假定算法是公开的，在这种情况下，专家才能对算法进行研究和测试，找出算法的安全漏洞。在算法公开的情况下，攻击者也可设计针对特定水印算法的攻击，一个典型的例子是针对回声隐藏算法的攻击，基本思想是：首先估计回声的延迟时间和衰减系数，然后使用检测结果将回声消除。另外一个典型例子是针对扩频音频水印的攻击方法，其基本方法是使用非线性滤波技术从含水印作品中获得水印信号的估计信号。首先从含水印音频作品 s' 中估计水印信号 w' ，同时使用含水印作品 s' 估计感知掩蔽阈值，使用感知掩蔽阈值对估计的水印信号 w' 整形并使用缩放因子缩放后获得最终的水印估计信号 w'' ，从含水印作品中 s' 中减去水印信号估计值 w'' 即获得原始载体信号的一个估计 c' 。

(3) 去同步攻击

去同步攻击的目的是使水印检测器和嵌入的水印无法对齐，从而使水印检测前的同步过程在计算上不可行。由于多数水印算法，尤其是基于相关检测的算法，要求检测前水印检测器需要与嵌入在作品中的水印完全对齐或者近似对齐。实施去同步攻击可采用延迟、时域缩放以及随机抽取时域采样值等方法。在音频信号中每秒钟随机抽取或者添加十几个采样值不会引起听觉失真，在音频信号相对平稳的区域可以抽取更多的采样值而不会引起失真。

2. 音频水印系统的鲁棒性评测

评价水印系统的性能指标主要有鲁棒性、透明性、载荷、计算复杂性等，根据应用的不同，一些指标会比另外一些更重要，其中鲁棒性和透明性是用于版权保护应用水印的基本要求，前面已经介绍了透明性（含水印音频的质量）的评价问题，这里主要介绍评价水印鲁棒性的已有标准。

既然水印系统面临如此众多的攻击方法，一个重要的问题是，水印算法的设计者和软件制造商如何测评所设计的算法是否达到特定应用环境的要求？水印技术的用户如何知道某个特定的算法是否符合他们的基本要求？当然软件制造商可以采用自己的测试标准，并将测试结果提供给用户，但是不同制造商采用不同的测试标准，测试结果之间没有可比性。另外一个方法是用户自己制定标准并组织测试，但是成本会很高，并且不是每个用户都有这种能力。解决这个问题的方法是采用一个统一的测试标准，标准由除了

算法提供者和使用者之外的第三方提供，第三方常常是相关行业的协会或者一个公开机构。这样，软件商在软件测试阶段可使用测试标准来确定该软件是否达到设计要求，用户可采用该测试标准对不同算法进行测试，来确定哪些算法满足自己特定应用的要求；另外，算法研究人员可采用测试标准来确定某个已有算法是否需要改进，或者对算法的改进是否的确提高了性能。国际唱片业协会（International Federation of Phonographic Industry, IFPI）在 1997 年提出了对音频数字水印系统的要求，安全数字音乐倡导者联盟（Secure Digital Music Initiative, SDMI）和日本音乐著作权协会（Japanese Society for Rights of Authors, Composers and Publishers, JASRAC）也提出了音频水印系统的鲁棒性标准，这三个标准属于行业协会提供的测试标准。另外一个重要的测评标准是 StirMark 中关于音频水印鲁棒性的标准，这部分工作主要由德国 Fraunhofer 研究所的研究人员进行，StirMark 中的音频测试系统属于开放系统。

（1）IFPI, SDMI 和 JASRAC 对音频系统鲁棒性的要求

IFPI 在 1997 年和美国唱片协会（The Recording Industry Association of America, RIAA）发起一项测试，其目的是寻找或确定一种在录音制品中嵌入不可感知信息的方法，包括在视频伴音或者其他多媒体伴音中。要求在重采样、转录以及数字/模拟转换等操作后仍能够提取出隐藏的信息，IFPI 要求参评水印系统载荷达到 20bps。

SDMI 是由 RIAA、IFIP 等在 1998 年发起成立的一个组织，SDMI 由 180 多家公司和组织构成，包括信息技术、消费电子、安全技术以及国际唱片业等行业和组织等。它的目标是制定一个开放的行业标准来保护数字音乐制品版权，从而有利于数字音乐市场的健康发展。出于同样的目的，JASRAC 也提出类似的测试标准 STEP2000 和 STEP2001。表 3.6 中给出了 STEP2000 和 SDMI 对音频水印鲁棒性的要求。

上述音频水印鲁棒性测试方法为不同算法间性能比较提供了基础，也为越来越多的算法研究人员采纳，在发表研究结果时，使用上述测评手段能提供具有可比性的结果。

（2）StirMark 中与音频鲁棒性评测有关的工作

StirMark 音频水印攻击基准与上述标准有所不同，上述标准是从水印技术用户的角度出发，对水印系统的性能的要求。StirMark 试图建立一个自动测评程序，例如建立自动测评服务器，用户通过网络向服务器提供算法的相关函数，由自动测评程序模拟各种攻击，然后将测评结果存储在数据库中供受测试方查询或者供其他水印研究人员参考。

表 3.6 SDMI 和 STEP2000 对音频鲁棒性的测评标准

信号处理操作	要 求										
数字到模拟转换	数字到模拟转换后进行模拟到数字转换										
均衡	采用具有如下特性的十段图示均衡器										
	频率（Hz）	31	62	125	250	500	1k	2k	4k	8k	16k
	增益（dB）	-6	+6	-6	+3	-6	+6	-6	+6	-6	+6
带通滤波	通带范围：100Hz～6kHz，过渡带衰减 12dB/oct										
时域拉伸和基音改变	+/-10%压缩和扩展										
编解码器 （使用典型数据率）	AAC, MPEG-4 AAC, MPEG-1 Audio layer 3, Q-Design, Windows Media Audio, Twin-VQ, ATRAC-3, Dolby Digital AC-3, ePAC, RealAudio, FM, AM, PCM										
添加噪声	添加白噪声，使信噪比为 40dB										
时间标度修改	沿时间轴进行+/-4%的缩放，不改变基音										
频率颤动	频率从 0～250Hz，方均根值为 0.5%										

续表

信号处理操作	要 求
添加回声	最大延迟 100ms, 最大衰减系数 0.5
解混响和环绕声处理	将立体声转换为单声道, 6 声道转换为立体声, SRS 音响还原, spatializer, Dolby surround 等环绕立体声处理
采样率改变	44.1~16kHz, 48~44.1kHz, 96~48/44.1kHz
动态范围改变	阈值选为 50dB, 最大压缩率 16dB, 攻击时间 10ms, 恢复时间 3ms
幅度压缩	16 比特转换为 8 比特

针对互联网应用环境, Fraunhofer 研究所的研究人员进一步探讨了如何采用 StirMark 对各种音频信号压缩攻击的效果进行自动评价。前面讨论了水印技术与多媒体有损压缩技术之间的矛盾, 这种矛盾在互联网环境下尤为突出, 并且可以预料随着音频压缩算法的发展, 音频水印会面临更大的威胁。目前互联网上使用的音频压缩算法主要包括: MP3 格式、VQF 格式和 WMA 格式三种, 其中在高速率情况下 MP3 和 VQF 感知性能较好, 在低速率情况下 WMA 格式的感知性能较好。实验表明, 不同算法在不同压缩率下对水印的攻击强度不同。Steinebach 等人试图建立一个不依赖于具体压缩算法的音频感知压缩模型, 并将该模型加入到 StirMark 标准中。该模型原理如图 3.24 所示。有损压缩仿真模型的基本思想是采用一系列信号处理操作来模拟有损压缩的效果, 包括前面讨论的滤波、动态范围改变和均衡等。该模型首先分析含水印待测试信号的特征, 从外部输入端获取压缩算法的类型以及压缩率等参数, 有损压缩仿真模型采用相应参数的信号处理操作序列来模拟压缩效果, 图 3.24 中的随机噪声用来模拟压缩算法的不可预测部分。

在另外一些应用场景中需要音频水印技术能够抵抗 AD/DA 转换以及噪声环境下放音录音等攻击, 这些应用场景包括在影片伴音中添加水印, 这样盗版者使用数字视频录像机录制的影片中将包含该电影院的标识码, 另外还可用于便携式音频监控器。在电视播放的歌曲中嵌入水印, 具有水印检测器的儿童玩具能够“听懂”歌曲, 并作出相应的动作。为了评价水印算法在上述应用场景下的鲁棒性, 需要将自动测评方法加入到 StirMark 中。Steinebach 等人考虑的仿真模型包括以下效果: ① 使用声卡进行 DA 转换。② 对信号进行音频放大, 并引入噪声。③ 考虑喇叭的频率响应的影响。④ 周围环境的影响, 例如房间的混响。⑤ 麦克风的频率响应特性。⑥ 麦克风信号放大作用。⑦ 用声卡进行 AD 转换。根据上述效果, AD/DA 仿真模型包括了一系列的信号处理操作, 包括: ① 添加噪声, ② 量化操作, ③ 修改信号的能量, ④ 滤波操作, 仿真硬件设备的频率响应特性。

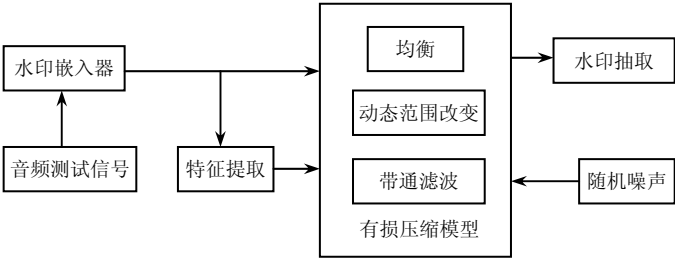


图 3.24 有损压缩仿真模型

从 StirMark 的有损压缩算法模型和 DA/AD 转换模型可见, 其基本思想是采用基本的信号处理操作的组合来模拟各种应用场景下的攻击, 对于用户而言, 不需要建立相应的测试环境, 只需要使用仿真软件就可以完成特定应用场景下水印系统鲁棒性能的测

试, 降低了测试成本, 减少了开发周期。

3.6.5 典型时域数字音频水印算法

回声隐藏和幅度调制是两种典型的时域数字音频水印算法。回声隐藏已经在隐写术中介绍过, 所以这里我们只介绍幅度调制方法。幅度调制的基本思想^[37]是通过改变两个或三个音频数据块的能量来嵌入水印。对每个长度为 N 的数据块, 其“能量”定义如下

$$E = \sum_{i=1}^N |c(i)| \quad (3.34)$$

其中, $c(i)$ 表示时域采样值。当信号的幅度较大时, 数据块的“能量”较高。假定使用两个连续数据块来嵌入水印, 通过修改每个数据块幅度, 可使数据块 A 和 B 的能量相同或不同。令 E_A 和 E_B 分别表示数据块 A 和 B 的能量。例如, 若 $E_A \geq E_B + \tau$, 则判定这一段音频信号中已经嵌入水印信息“0”; 若 $E_A \leq E_B - \tau$, 则判定这一段音频信号中已经嵌入了水印信息“1”, 如果以上两种情况都不成立, 则判定这一段音频信号没有水印嵌入。然而, 这种方法存在严重问题。假定数据块 A 中的能量比数据块 B 高得多, 并且要嵌入的水印信息是“0”, 那么没有任何问题。在其他情况下, 需要将 E_A 变得大于 E_B 。只要相邻数据块的能量差异很大, 修改造成的效果就会很明显并且听起来很不自然, 这种方法能够将强音乐段变成轻音乐段或者相反。

通过使用三个或者多个数据块, 上述问题能缓解。使用多个数据块, 通过将负担分布到多个数据块中来减轻人工修改造成的不自然效果。下面介绍采用三个相邻数据块能量关系来嵌入水印的算法。设 E_1 、 E_2 、 E_3 表示三个相邻数据块的短时能量, 将三个数据块按能量大小排序, 设排序后的能量为 E_{\max} 、 E_{mid} 、 E_{\min} , 则能量块间的差异可计算为

$$\Delta E_1 = E_{\max} - E_{\text{mid}}; \quad \Delta E_2 = E_{\text{mid}} - E_{\min} \quad (3.35)$$

嵌入算法通过修改能量块间差异的差值来嵌入信息具体如下。① 当嵌入的比特位为“1”时, 如果 $\Delta E_1 - \Delta E_2 \geq (E_{\max} + 2E_{\text{mid}} + E_{\min}) \cdot d$, 则不对信号作修改, 否则增加 E_{\max} 或者减小 E_{mid} , 直到满足条件 $\Delta E_1 - \Delta E_2 \geq (E_{\max} + 2E_{\text{mid}} + E_{\min}) \cdot d$; ② 当嵌入的比特位为“0”时, 如果 $\Delta E_2 - \Delta E_1 \geq (E_{\max} + 2E_{\text{mid}} + E_{\min}) \cdot d$, 则不对信号作修改, 否则增加 E_{mid} 或者减小 E_{\min} , 直到满足条件 $\Delta E_2 - \Delta E_1 \geq (E_{\max} + 2E_{\text{mid}} + E_{\min}) \cdot d$ 。其中参数 d 控制修改的强度。水印提取算法采用盲检测方法, 设同步过程已经完成, 相邻三个数据块的短时能量分别为 E_1' 、 E_2' 、 E_3' , 排序后表示为 E_{\max}' 、 E_{mid}' 、 E_{\min}' , 则计算

$$\Delta E_1' = E_{\max}' - E_{\text{mid}}'; \quad \Delta E_2' = E_{\text{mid}}' - E_{\min}' \quad (3.36)$$

如果 $\Delta E_1' \geq \Delta E_2'$, 则提取的比特位为“1”, 否则提取的比特位为“0”。

为达到水印的不可感知性, 该算法采取两个措施: ① 使时域修改量缓慢变化, ② 采用频域掩蔽模型, 通过修改参数 d 来保证水印信号的频谱位于全局掩蔽阈值之下。

3.6.6 典型变换域数字音频水印算法

音频水印技术中常使用的变换域算法包括傅里叶变换域、离散余弦变换域、倒谱域和小波域等。这里介绍一种采用扩频思想在数字音频信号的离散余弦变换域嵌入视觉可辨水印的方法^[38]。嵌入的水印是一幅二值图像, 例如商标图案等。

1. 水印嵌入算法

在数字音频 c 中嵌入二值图像 m 的框图如图 3.25 所示。水印嵌入过程如下。

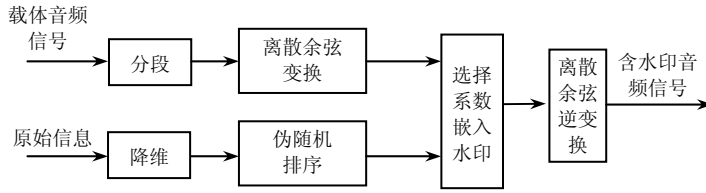


图 3.25 DCT 域数字音频水印嵌入算法框图

(1) 划分数据段：在数字音频信号 $c=\{c(i)\}$ 中划分出 $A \times B$ 个互不相交的、长度为 N 的数据段，它们可表示为

$$c = \{c^{(l)}, 0 \leq l < (A \times B)\} \quad (3.37)$$

其中， $c^{(l)}$ 表示第 l 个音频数据段，它可表示为

$$c^{(l)} = \{c(l \times N + i), 0 \leq i < N\} \quad (3.38)$$

(2) 原始信息降维：由于嵌入的信息是二维图像 m ，要将其嵌入到一维数字音频信号中，必须进行降维处理，将其转化成一维序列 v ，即

$$v = \{v(l) = m(i, j), 0 \leq i < A, 0 \leq j < B, l = i \times B + j\} \quad (3.39)$$

通过降维操作，图像中的像素 $m(i, j)$ 用序列 v 中第 l 个元素 $w(l)$ 表示。

(3) 水印生成（伪随机排序）：为消除序列 v 中相邻元素的相关性，提高嵌入水印的鲁棒性，采用线性反馈移位寄存器生成的伪随机序列对 v 中的所有元素进行排序，即

$$w = \text{Permutation}(v) = \{w(l) = v(l'), 0 \leq l, l' < (A \times B)\} \quad (3.40)$$

通过排序操作，序列 v 中的第 l' 个元素移动到第 l 个元素的位置上。

(4) 离散余弦变换：对所有的音频数据段分别进行离散余弦变换，则

$$C^{(l)} = \text{DCT}(c^{(l)}) \quad (3.41)$$

其中 $C^{(l)} = \{C_i^{(l)}, 0 \leq i < N\}$ ， $C_i^{(l)}$ 是第 l 个音频段 $c^{(l)}$ 的离散余弦变换 $C^{(l)}$ 中的第 i 个系数。

(5) 选择中频系数：在离散余弦变换域内确定数字音频信号的中频系数，即在 $C^{(l)}$ 内选取第 p 个系数 $C_p^{(l)}$ 作为中频系数用于嵌入序列 w 中相应的元素 $w(l)$ 。数字音频数据段 $c^{(l)}$ 的数据个数为 N ，其离散余弦变换结果 $C^{(l)}$ 中也含有 N 个 DCT 系数。其中第 0 个 DCT 系数 $C_0^{(l)}$ 为直流分量，其他的 $N-1$ 个 DCT 系数是由低频到高频的交流分量。为了提高嵌入水印的鲁棒性，通常选取频率较低的交流分量（ $C_1^{(l)}$ 除外）作为中频系数，如选取 $C_2^{(l)}$ 作为中频系数（ $p=2$ ）。

(6) 嵌入水印：修改中频系数 $C_p^{(l)}$ 嵌入序列 w 中元素 $w(l)$ ，即

$$S_i^{(l)} = \begin{cases} C_i^{(l)}(1 + \alpha w(l)) & \text{若 } i = p \\ C_i^{(l)} & \text{否则} \end{cases} \quad (3.42)$$

其中， α 是比例系数，用于控制修改量。若 α 的取值过小，则嵌入水印的鲁棒性比较差；若 α 的取值过大，则会降低数字音频信号的使用价值，引起听觉失真。比例系数 α 的取值范围应根据音频信号的应用背景确定。

(7) 离散余弦逆变换：对 $S^{(l)}$ 的所有数据段分别进行离散余弦逆变换，即得到各段的含水印音频段 $s^{(l)}$

$$s^{(l)} = \text{IDCT}(S^{(l)}) \quad (3.43)$$

2. 水印抽取算法

这部分讨论的数字音频信号的水印嵌入算法是源抽取算法，从含水印数字音频信号 s' 中抽取水印的框图如图 3.26 所示。数字音频信号的水印抽取过程及表示式如下。

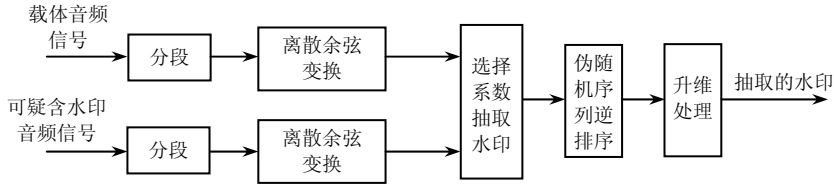


图 3.26 DCT 域数字音频信号的水印抽取算法框图

(1) 划分数据段：对原始音频 c 和待检测音频 s' 分别进行分段处理，即

$$c = \{c^{(l)}, 0 \leq l < (A \times B)\} \quad (3.44)$$

$$s' = \{s'^{(l)}, 0 \leq l < (A \times B)\} \quad (3.45)$$

(2) 离散余弦变换：对 c 和 s' 分别进行分段离散余弦变换，则有

$$C^{(l)} = \text{DCT}(c^{(l)}) \quad (3.46)$$

$$S'^{(l)} = \text{DCT}(s'^{(l)}) \quad (3.47)$$

(3) 抽取水印：

$$w'(l) = \frac{1}{\alpha \times C_p^{(l)}} (S_p^{(l)} - C_p^{(l)}) \quad (3.48)$$

由于抽取的信息是实数型的，必须将 $w'(l)$ 归一化成二值形式，即

$$w'(l) = \begin{cases} 1 & |w'(l)| > \text{Threshold} \\ 0 & \text{Otherwise} \end{cases} \quad (3.49)$$

其中 Threshold 是由用户确定的门限值。

(4) 伪随机逆排序：对 w' 进行伪随机序列逆排序得序列 v' ，即

$$v' = \text{InversePermutation}(w') = \{v'(l) = w'(l'), 0 \leq l, l' < A \times B\} \quad (3.50)$$

(5) 恢复信息：对 v' 进行升维处理，即将一维的序列转换成二维的图像输出 m'

$$m' = \{m'(l) = v'(i, j), 0 \leq i < A, 0 \leq j < B, l = i \times B + j\} \quad (3.51)$$

3. 仿真实验

在仿真实验中，原始信号采用 8 位 22.05kHz 采样/秒的数字音频信号，其波形如图 3.27 (a) 所示。数字水印采用 64×64 的二值图像，如图 3.28 (a) 所示。在水印的嵌入过程中，音频数据段的长度 $N=8$ ，比例系数 $\alpha=0.1$ ，选取每个音频数据段的第 2 个 DCT 系数 ($C_2^{(l)}$) 作为中频系数。图 3.27 (b) 是含水印数字音频信号的波形图 (SNR=45.72dB)。

对数字音频信号攻击的方式主要有滤波、有损压缩等。图 3.28 (b) 是从经过 5:1 压缩处理的数字音频信号中抽取的水印 (归一化相关系数 NC=0.74)；图 3.28 (c) 是从经过低通滤波处理的版本中抽取的水印 (NC=0.82)；图 3.28 (d) 是从经过平均值滤波处理的版本中抽取的水印 (NC=0.83)；图 3.28 (e) 是从未经任何处理的版本中抽取的水印 (NC=1.00)。从仿真实验结果可以看出，抽取的水印和原始水印具有明显的视觉相似性，因此很容易对数字音频信号的版权归属问题作出结论；这部分讨论的基于分段离散余弦变换的水印嵌入算法对压缩、滤波等操作具有较强的鲁棒性。

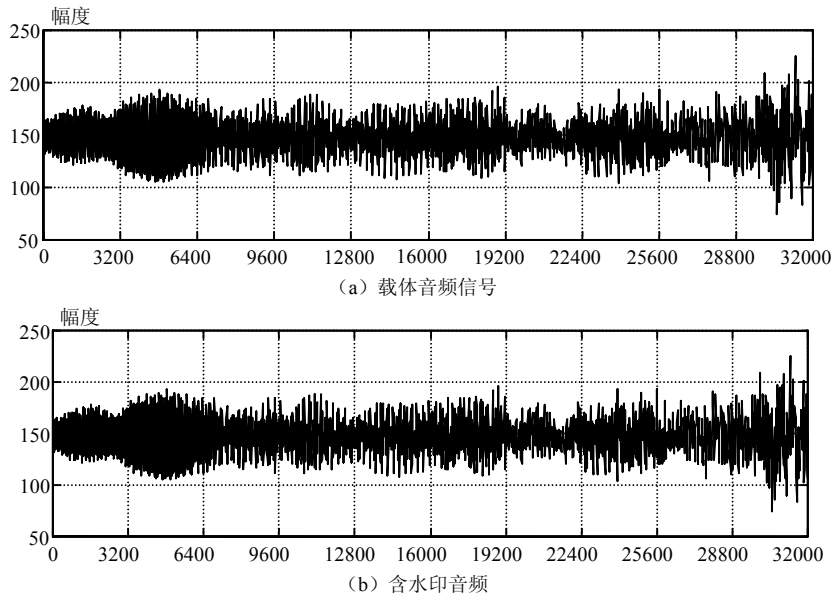


图 3.27 数字音频信号波形图



图 3.28 原始的水印和抽取的水印

3.6.7 典型压缩域数字音频水印算法

在互联网上, 音频制品特别是音乐作品多以压缩格式存在, 所以有必要研究如何直接在压缩后的比特流中嵌入与载体作品相关的信息。另外, 设计水印系统时往往需要考虑如何抵抗感知编码的影响, 压缩域水印方法将水印嵌入过程和感知编码过程相结合, 在压缩编码的同时嵌入了水印, 从而避免了感知编码对水印的攻击。压缩域水印和比特流水印的弱点是抵抗格式转换的鲁棒性差, 所以一般适用于对音频内容检索、标注等应用。本小节介绍 MPEG-2 AAC 压缩域水印和 MPEG-1 Audio Layer II 比特流水印技术。

1. MPEG-2 AAC 压缩域水印算法

在 MPEG-2 音频部分的标准中包含了与 MPEG-1 前后向兼容和不兼容的多通道音频编码标准。使用与 MPEG-1 前后向兼容的标准能够使 MPEG-1 解码器解码 MPEG-2 压缩的多通道音频信号, 同时使 MPEG-2 解码器解码由 MPEG-1 压缩的双声道音频信号, 但是后向兼容带来的缺点是使原来被掩蔽的噪声信号变得可以听见。MPEG-2 AAC (Advanced Audio Coding) 是 MPEG-2 中非后向兼容的音频压缩标准, 支持多通道环绕立体声编码, 并且不存在后向兼容标准中的噪声问题。Lacy 等人提出了一种 MPEG-2 AAC 压缩域水印^[39], 这种方法通过修改尺度因子 (Scale Factor)、谱线量化值以及编码用的霍夫曼编码表来嵌入信息。在 MPEG AAC 中, 谱线被分为 49 个尺度因子带 (Scale Factor Band, SFB), 每个带包含 4~32 根谱线。每个 SFB 拥有 1 个尺度因子 (用来设定量化

器阶矩)、1 个霍夫曼表 (AAC 标准规定了 12 个表, 其中一个空的, 能用来嵌水印的 SFB 的霍夫曼表不能是空的)。每根谱线的系数用一个整数值 (量化值) 来表示。设 $A=\{f_i, H_i, \{q_{ij}\}, i \in \Gamma\}$ 表示一个三元组, 其中 f_i 表示尺度因子, H_i 表示霍夫曼编码表, $\{q_{ij}\}$ 表示量化后的频谱系数, Γ 是所有嵌入水印的 SFB 序号集合 (这些 SFB 是事先基于感知模型选择好的)。水印嵌入算法首先选择一系列因子 $\{x_i=2^{n_i}, i \in \Gamma\}$, 然后使用各因子 $\{x_i\}$ 去除各尺度因子 $\{f_i\}$, 同时使用各 $\{x_i\}$ 乘以对应 SFB 的所有量化后的频谱系数 $\{q_{ij}\}$, 并将水印标记 $\{w_{ij}\}$ 加到那些非零的修改后的频谱系数上。设 $A'=\{f'_i, H'_i, \{q'_{ij}\}\}$ 表示修改后的三元组, 则嵌入算法可以表示为

$$\begin{cases} \{f'_i, H'_i, \{q'_{ij}\}\} = \{f_i, H_i, \{q_{ij}\}\} & i \notin \Gamma \\ \{f'_i, H'_i, \{q'_{ij}\}\} = \{f_i/x_i, H''_i, \{q_{ij} \cdot x_i + w_{ij}\}\} & i \in \Gamma \end{cases} \quad (3.52)$$

其中, H''_i 表示能够编码最大 $q_{ij} \cdot x_i + w_{ij}$ 值的最小霍夫曼编码表。

2. MPEG-1 Audio Layer II 比特流水印算法

Qiao 等人提出通过修改 MPEG-1 Audio Layer II 比特流中的尺度因子和编码后的采样值来嵌入水印^[40]。MPEG-1 Audio Layer II 的帧结构如图 3.29 所示。数据帧是 MPEG 音频中最小的解码单位, 它包括音频数据、帧头、CRC 校验码以及辅助数据。在 MPEG-1 音频编码第二层的数据帧中, 每帧包含 1152 个样值, 每 12 个样值组成一组, 每个子带包含 3 组。对不同组的样值, 编码器可以采用不同的尺度因子, 或者对几个组使用相同的尺度因子。MPEG-1 Audio Layer II 的数据帧的载包含三部分: ① SFSI-尺度因子选择信息, 包含各个组对尺度因子的选择信息。② SF-尺度因子, 包含本帧所使用的所有尺度因子。③ Sample-包含本帧所有编码后的样值。

帧头	CRC	比特分配	尺度因子	编码后的样本	辅助数据
----	-----	------	------	--------	------

图 3.29 MPEG-1 Audio layer II 的帧结构

在水印信号的生成过程中, 首先采用标准加密算法 DES 对每帧 MPEG-1 音频数据帧加密, 设 $c^{(i)}$ 表示第 i 帧 MPEG-1 数据帧, 使用密钥 k 产生随机字节序列 RBS。

$$\text{RBS} = \{\text{RBS}_i \mid \text{RBS}_i = \text{DES}(c^{(i)}, k)\} \quad (3.53)$$

其中, RBS_i 为 RBS 的第 i 个字节, w_i 表示水印位流的第 i 位, 则水印位的产生方法为

$$w_i = \begin{cases} -1 & \text{若 } \text{RBS}_i \text{ 为偶数} \\ 1 & \text{否则} \end{cases} \quad (3.54)$$

(1) 在尺度因子中嵌入水印

尺度因子是一个乘子, 它能够使编码器充分利用量化器的量化范围, 解码器将尺度因子乘以解码后的样值来重建音频信号。在 MPEG-1 Audio Layer II 中采用 6 比特来表示每个尺度因子, 具有 63 个量化级。实验表明, 将尺度因子改变一个量化级, 人类听觉系统一般不能感知到差别。在尺度因子中嵌入水印的方法直接将水印位加到尺度因子索引上, 使尺度因子改变一个量化级。嵌入算法如下: 设 $f_i(j)$ 为第 i 个尺度因子, 其索引值为 j ($0 \leq j \leq 62$), w_i 表示第 i 个水印位值, $f'_i(j)$ 表示嵌入水印后的尺度因子, 则

$$f'_i(j) = \begin{cases} f_i(j) & \text{若 } j + w_i = -1 \text{ 或 } 63 \\ f_i(j + w_i) & \text{否则} \end{cases} \quad (3.55)$$

该方法的缺点是嵌入容量不高, 因为每帧数据只有很少的尺度因子。另外, 使用该

方法不能嵌入多个水印, 否则会使尺度因子变化多个量化步长而引起感知失真。

(2) 在编码后的样本中嵌入水印

修改样本的基本方法是將水印位加到解码后的样本序列上, 但这样会引起可感知失真。一个改进方法是每隔若干样本, 修改其中的一个样本。实验表明, 这种修改方法引入的噪声一般不可感知。水印产生过程需要作如下修改: 设 N 表示间隔的样本数, 则

$$w_i = \begin{cases} -1 & \text{若 } \text{RBS}_i = 0 \pmod{N} \\ 1 & \text{若 } \text{RBS}_i = 1 \pmod{N} \\ 0 & \text{否则} \end{cases} \quad (3.56)$$

水印嵌入过程与使用尺度因子嵌入水印类似, 但要保证嵌入水印后的样本不会出现比特位全“1”的情况, 因为这种情况在 MPEG 音频标准中是不合法的。设 c_i 表示第 i 个样本, s_i 表示嵌入水印后的样本, 则嵌入算法表示为

$$s_i = \begin{cases} c_i & \text{如果 } (c_i + w_i) \text{ 的每个比特都是 } 1 \\ c_i + w_i & \text{否则} \end{cases} \quad (3.57)$$

上述两种方法实际上都采用了扩频思想。

3.7 数字视频水印技术

视频水印就是加载在数字视频上的数字水印, 它利用视频数据中普遍存在的冗余数据和随机性把表征版权的信息嵌入到原始视频中, 从而保护数字产品版权或完整性, 确保版权所有者的合法权益。视频水印技术的出现最初是为了保护数字视频产品 (如 VCD、DVD、VOD 等) 的版权, 但因为其具有不可感知性、鲁棒性和安全性等特点, 近年来其应用领域得到不断的扩展, 潜在应用包括版权保护、广播监控、复制控制、内容认证。相对于图像水印, 学者们对基于视频产品的数字水印技术研究较少。但是, 随着多媒体技术的发展, 视频产品越来越多 (例如 DVD、CD、多媒体教材、录像带等), 对于视频产品的版权保护亟须解决, 因此国内外越来越多的学者逐渐投入到视频水印的研究当中。本节首先介绍视频数字水印的特点和面临的挑战, 然后介绍视频数字水印系统模型和算法分类, 最后概述介绍常见的一些视频水印算法, 包括原始域和压缩域算法。

3.7.1 数字视频水印技术的特点和面临的挑战

1. 视频水印的主要特征

由于数字视频是连续播放的图像序列, 其相邻帧之间的内容有高度的相关性, 连续帧之间存在大量的数据冗余, 使得视频水印容易遭受帧平均、帧丢弃、帧交换等各种攻击, 而且目前为了节约视频数据存储空间和便于传输, 视频的主要存在模式是压缩格式的, 视频水印在很大程度上是与压缩编码标准紧密联系在一起的, 因此视频水印除了具有一般水印技术的特征外, 还有一些特殊的要求, 视频水印的特征可以概括为以下几点。

(1) 稳健性

视频水印应该能够抵抗各种无意或故意的攻击, 包括帧平均、帧丢弃、帧交换等专门针对视频水印的攻击。

(2) 不可见性

视频嵌入水印后不会影响视频质量, 从而确保视频数据的商业价值。

(3) 安全性

视频水印中的信息应是安全的, 难以被篡改或伪造, 未经授权的用户无法正确地检测、提取或移除水印。

(4) 视频速率的恒定性

水印嵌入视频数据后不能改变视频流码率, 须服从传输信道规定的带宽限制, 否则将有可能造成解码后的视频图像和声音的失步, 降低视频质量。

(5) 与视频编码标准相结合

对于压缩视频, 水印设计必须与视频编码标准结合, 对于在原始视频中嵌入的水印也必须考虑编码标准, 否则嵌入的水印有可能在编码中消失。

(6) 实时性

水印的嵌入和检测提取算法复杂度不能高, 必须在短时间内完成, 以保证视频数据的实时编解码。

(7) 盲检测

水印检测原则上不能使用原始视频, 以确保水印检测能够实时完成。

(8) 水印容量

嵌入的水印必须能够携带足够多的信息。对于视频水印, 规定水印容量为单位时间嵌入水印信息的数据量, 通常要求水印算法有尽可能高的嵌入水印速率。

2. 视频水印技术面临的挑战

由于视频水印的一些特殊性质, 使得视频水印算法设计与静止图像水印算法设计之间存在许多差异, 现有的图像水印算法还不能很好地保护视频数据, 视频水印技术面临着以下一些新的挑战。

1) 视频水印算法要有更强的健壮性

由于数字视频是由连续播放的视频帧所构成, 相邻帧之间的内容具有高度的相关性, 导致视频水印除了可能遭到一般对图像水印的攻击形式外, 还可能遭到一些针对视频水印的特殊攻击形式(如帧重组、帧删除、帧间统计平均和统计共谋等)的出现, 因此要求视频水印技术要具有抵抗这些攻击的能力。针对视频水印的主要攻击和处理有:

(1) 无恶意的处理。主要指数字视频的使用者为了更好地管理或使用视频数据资源而对其视频数据进行的各种处理, 包括视频在传输过程中引入噪声、DA/AD 转换引起的信号失真、传输编码或视频格式的转换、空间分辨率(NTSC、PAL、SECAM)之间的切换、视频帧率的改变, 以及帧重组、帧插入、帧删除、淡入淡出视觉特技等。

(2) 故意的攻击。故意攻击一般可分为简单攻击、检测失效攻击、混淆攻击和移除水印攻击四类, 视频水印的攻击也基本由这四类构成。其中对于单个的视频帧, 基于静态图像的攻击方法一般有效。而对于连续的视频帧, 恶意用户更容易利用视频帧连续且相邻帧高度相关的特点, 进行统计平均和统计共谋攻击。

2) 视频水印算法要具备实时性

视频水印算法在多数情况下有实时性的要求, 这也是视频水印算法与静止图像水印算法之间的重要区别。一般而言, 对于在静止图像中嵌入或提取水印, 几秒或更长的时间是可以接受的, 但对于视频数据而言, 由于视频帧的帧率较高(如 25 帧/秒), 以确保视频数据流的平滑性, 所以较长的水印嵌入或提取的延迟时间会降低视频帧率, 从而严

重影响视频的质量。视频数字水印技术对实时性有多层次的要求：一种要求嵌入和提取都必须实时，称为强实时；另一种没有强实时要求，或要求嵌入时满足实时性，或要求提取时满足实时性，称为弱实时。对于在数字电视广播中应用数字水印，往往要求实时性较高，因而是强实时的；另外一种特殊应用，如广播监视，水印检测器必须能够实时地检测水印，而在水印嵌入过程中对实时性的要求不是很高，这种应用是弱实时的。

3) 视频水印算法要具有盲检测性和随机检测性

视频水印的检测原则上不能使用原始视频数据，这是因为在检测时使用原始视频数据会大大增加运算的复杂度，使得水印算法无法实现实时性要求，另外视频信号载体的数据量较大，如果采用非盲检测技术，检测时需要原始视频数据做参考也是不现实的。视频水印的随机检测性指可以在视频的任何位置，短时间内检测出水印。在许多实际的视频水印应用当中，不可能从视频的开始位置按播放顺序一步步地检测出水印，而且嵌入水印的视频也可能遭受帧删除、帧重组等攻击，因此视频水印技术要保证能够在视频的任何一个位置，在一小段视频图像序列中能够检测到水印。

4) 视频水印算法要与视频编码标准相结合

视频数据由于其数据量很大，在存储、传输中通常先对其进行压缩编码。如果在压缩视频流嵌入水印，很显然要与视频压缩编码标准相结合；如果是在原始视频中嵌入水印，由于水印嵌入是利用视频的冗余数据来携带信息，而视频压缩编码则需要去除视频中的冗余数据。如果不考虑视频压缩编码标准而盲目地嵌入水印，则嵌入的水印很可能在编码过程中完全或大部分丢失。

3.7.2 视频数字水印系统的模型和算法分类

1. 视频数字水印系统的模型

视频水印技术是在静止图像水印技术的基础上逐渐发展起来的，最初视频水印是将视频看作一个个单独的帧构成的图像序列，再运用图像水印的方法嵌入水印。这种方法的缺点是没有考虑到视频在短时间内帧内容高度相关的特性，水印很容易被帧平均的方法去除。现在已经有许多针对视频水印不同应用而提出的视频水印算法。由于数字视频编解码系统与静止图像编解码的不同，视频水印的嵌入和提取过程和图像水印的嵌入提取过程有很大的不同，视频水印的算法根据嵌入策略可以分为在未压缩的原始视频图像、视频编解码器、压缩后的视频码流中嵌入水印三种方案，图 3.30 显示了视频水印模型的几种嵌入和提取方案^[33]。

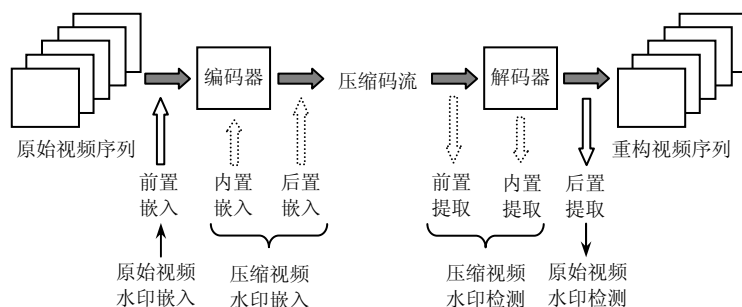


图 3.30 视频的水印嵌入和提取模型

(1) 方案一：前置嵌入

水印直接嵌入到未经过编码的原始视频图像序列中，然后再对含有水印信息的视频图像进行编码压缩。这种方案可以充分利用静止图像的水印技术，结合视频帧的结构特点，形成适用于视频水印的方案。这类算法的优点是水印算法比较成熟，静止图像水印的许多思想方法，如扩频、人类视觉模型、图像自适应、水印不可逆、同步检测机制等都可以推广应用到视频水印系统中。但这种方案也有明显的缺点，即会增加视频码流的数据比特率，影响视频速率的恒定性；嵌入水印后的视频数据经压缩编码后有可能丢失水印；对于已压缩的视频，需要先进行解码，然后嵌入水印后再重新编码，增加了计算的复杂性并降低视频的质量。

(2) 方案二：内置嵌入

该方案在编码压缩时嵌入水印。当今视频压缩的标准包括 ISO/ IEC 的 MPEG-1、MPEG-2、MPEG-4 和 ITU-T 的 H.261、H.263 等，它们的基本编码思想是运动补偿预测和基于块的变换编码。在编码压缩时嵌入水印，可以直接与视频编码器相结合，通过利用视频数据压缩的原理，如去空域冗余的变换、量化和熵编码技术，去时域冗余的运动补偿、运动表示、运动估计技术，利用编码数据的特性，水印的嵌入和提取处理可以比较简单，能够实现水印嵌入和提取的实时处理。这种方案的水印嵌入过程比较简单，水印一般嵌入在变换域系数中，不会增加视频流的数据比特率；另外，由于其是将水印嵌入在变换域中，并和编码过程结合紧密，可以设计出抵抗多种攻击的水印算法。但该方案需要修改编码器和解码器，而且存在 GOP 的误差积累。

(3) 方案三：后置嵌入

在压缩域中嵌入水印，即水印直接嵌入到编码压缩后的比特流中。该方案的显著优点是没有解码和再编码过程，因而不会造成视频质量下降，同时计算复杂度较低。其缺点是由于压缩比特率的限制而限制了嵌入水印数据量的大小，嵌入水印的强度受视频解码误差的约束，嵌入策略受相应视频压缩算法和编码标准的局限。例如，可通过修改视频流中的可变长编码（VLC）来隐藏水印信息，这样可充分利用视频压缩编码标准，不需对压缩视频流完全解码再编码，计算复杂度小，嵌入水印的速率相对较高，但其缺点是对信道干扰和视频处理的抵抗能力较差，按同样算法在可标记的 VLC 码幅度值的最不重要位上加入随机比特就可破坏水印，传统的滤波、重采样和时域缩放等处理也会影响水印地提取。目前也有些算法提出在运动矢量中嵌入水印，将水印嵌入在幅度值大且相角变换小的运动矢量中，在压缩视频序列中，大部分的帧是运动补偿编码帧，所以在运动矢量中隐藏水印信息可以更加有效地利用视频比特流中的信息。

2. 数字视频水印算法的分类

视频数字水印的分类算法很多，也没有统一的标准，大体分类如下所示。

(1) 按载体类型分类

按载体类型分类，可以分为基于原始视频算法和基于压缩视频方法。基于原始视频的水印算法，是对未经编码的视频流数据直接进行处理，在原始视频数据中嵌入水印；基于压缩视频的水印算法，则与某种压缩标准相结合（如 MPEG-2 或 MPEG-4），在编码视频流中嵌入水印。

(2) 按嵌入域分类

按嵌入域分类，主要分为空间域（或时域）算法和变换域（频率域）算法。空间域算法是用待嵌入的信息替换载体信息的冗余部分。一种简单的替换算法就是用待嵌入信

息位替换载体中的一些最低有效位,只有知道隐藏信息嵌入的位置才能提取信息。变换域算法是在载体视频的某个变换域嵌入水印,如离散傅里叶变换(DFT)、离散余弦变换(DCT)、小波变换域(DWT)、离散哈达玛变换(DHT)等。变换域算法有很多优点,很多图像及视频水印算法都是基于变换域的,其主要优点有:①在变换域中嵌入的信号能量可以分布到空域的所有像素上;②在变换域中,人的感知系统的某些掩盖特性可以更方便地结合到编码过程中,有利于提高水印的鲁棒性;③变换域方法与大多数国际标准兼容,可直接实现压缩域内的算法,提高效率。

(3) 按密钥分类

若嵌入和提取密钥相同,称其为对称水印;否则称为非对称水印。

(4) 按检测和提取水印时是否需要原载体信号

按检测和提取水印时是否需要原载体信号分为盲水印算法(检测和提取水印时不需要载体信号)和非盲水印算法(检测和提取水印时需要载体信号)。由于视频的数据量非常大,对视频水印一般都采用盲水印算法;而对图像水印,可采用非盲水印算法。

(5) 按水印的特性分类

按水印的特性分类,可以分为鲁棒性水印和脆弱性水印。脆弱性水印对信号的改动比较敏感,主要用于篡改提示;鲁棒性水印要能够经受各种有意或无意的攻击。视频水印主要研究鲁棒性水印。

(6) 按水印能否反向求出

按水印能否反向求出分为可逆水印和非可逆水印。非可逆水印对基于估计水印的攻击具有很强的鲁棒性。

3.7.3 典型原始域视频水印算法

原始域视频水印是直接对未压缩的视频数据进行处理,与视频编码格式无关。因此,可以分为两种情况:①可以直接获得原始视频流数据。此时,可以直接在原始的视频流中完成水印的嵌入或提取,处理比较简单;②只能得到编码的视频流数据。此时,需要首先对编码视频进行解码,然后再嵌入或提取水印。在进行水印处理之后,如果有必要再重新压缩,其处理相对复杂。如果存在一些特殊的要求,比如要求嵌入水印前后的编码码流的长度保持不变,则处理更为复杂。下面介绍一种基于块分类的自适应视频水印处理算法^[41,42]。算法同时考虑帧内和帧间的信息,根据运动信息和区域复杂度对原始视频的图像块进行分类。在帧内,对于 8×8 的图像块按其是否包含细节信息(边缘或纹理)来进行分类;同时考虑人眼对于静止物体和运动物体不同的视觉特性,在相邻帧间进行运动检测,将图像块分为慢速运动区域和快速运动区域两类。通过这两层检测机制,选择既包含细节信息、又属于快速运动区域的图像块来嵌入水印,这样使得水印嵌入的位置自适应于人类视觉系统和视频信号的特性。此外,该算法克服了大多数自适应水印处理算法不能够实现盲提取的缺点,而且水印检测及提取过程中不需要参考其他附加同步信号。仿真实验验证了算法的有效性。

1. 嵌入区域的自适应选择

由于视频信号比静止图像多出一维空间,通常认为视频序列具有更大的空间来嵌入水印。但是大量实验表明,嵌入水印很容易引起视频质量的下降,而且还会产生在静止图像水印中不会出现的问题,例如闪烁。因此,视频水印对于不可见性的要求更为严格。

有关心理视觉的研究表明,人眼对于各种环境有不同的敏感度。例如,人眼对于纹理复杂区域中所产生的失真并不敏感;同时由于人类视觉的惰性,人眼对于高速运动的目标的细节部分也不是很敏感。视频序列与静止图像的不同在于它包含运动部分,具有变化的特性,而对于快速运动的物体人眼敏感度会有所下降,因此可以在运动区域嵌入较高强度的水印分量。所以,我们在相邻帧间利用运动检测器,将图像块划分成慢速运动(Slow-Motion)区域和快速运动(Fast-Motion)区域两类,然后将水印信息嵌入快速运动区域,从而使得水印嵌入位置自适应于视频序列中物体的运动。另一方面,根据人类视觉系统的特性,掩蔽阈值受到亮度、纹理复杂度等多种因素的影响。局部区域背景越亮,所包含的细节信息(纹理或边缘)越丰富,门限值就越高。因此,按照图像块包含细节信息的情况将其分为高细节(High-Detail)区域和低细节(Low-Detail)区域两类,然后选择高细节区域嵌入水印,使得水印嵌入位置自适应于视频图像的内容。

大量实验已经证明,将视频序列的每一帧分成 8×8 的图像块,人眼对于那些包含高细节信息并且沿时间轴变化较快的图像块中所产生的失真并不敏感,因此可以在这些图像块中嵌入水印。为了对所有的图像块进行分类,引入了两层检测机制:运动检测和细节分类。同时为了减小计算复杂度,利用 DCT 系数的能量作为分类依据。DCT 直流系数表示图像块的平均亮度,它反映该图像块的基本信息,因此可利用相邻帧间对应 DCT 直流系数的差值来检测当前帧内图像块内容有无变化,并将它们划分为慢速运动区域和快速运动区域。另一方面,当图像块中相邻像素点的灰度值发生较大变化时,信号的大部分能量都集中在 DCT 交流系数上,因此 DCT 交流系数能量可被用来确定图像块中是否含细节信息(纹理或边缘)。其他一些检测算法(如 Sobel 算子等)也可采用,但实验结果表明这里所给出的检测依据在分类准确性和计算复杂度两方面取得了较好的折中。

嵌入区域的自适应选择过程如图 3.31 所示。假设一段视频由 L 帧组成,每一帧的图像大小为 $A \times B$,令 c_i 表示原始视频中的第 i 帧,其中 $i=1, 2, \dots, L$ 。首先,将当前帧 c_i 分割成互不重叠的大小为 8×8 的图像块 c_{ij} ,这里 c_{ij} 表示第 i 帧的第 j 个图像块, $j=1, 2, \dots, [A \times B / 64 - 1]$ 。对每一个图像块进行 DCT 变换,得到 DCT 系数块 C_{ij} 。将 DCT 系数按“之”字形顺序排列,其中 $C_{ij,0}$ 表示直流系数。然后图像块的分类过程分两步进行。第一步,进行帧内细节信息检测,根据 DCT 交流系数能量的大小对图像块进行分类。这里, DCT 交流系数的能量用 $E_{AC}(i, j)$ 表示,即

$$E_{AC}(i, j) = \log \left(\sum_{l=1}^{63} (C_{i,j,l})^2 \right) \quad (3.58)$$

式中的对数运算是为了缩小取值范围,并保持单调性。对每一个 DCT 系数块 C_{ij} ,如果 $E_{AC}(i, j)$ 小于预先给定的阈值 T_D ,那么相应的图像块被划分为低细节区域;否则,相应图像块被划分为高细节区域,这一类图像块集合表示为 $X_i = \{c_{i,d_1}, c_{i,d_2}, \dots, c_{i,d_p}\}$,其中 d_j 表示高细节图像块的索引, $1 \leq j \leq p$, p 为高细节块总数。第二步,进行帧间运动检测,将当前帧 DCT 系数块与相邻帧中对应 DCT 系数块进行比较。简单起见,这里只参考前一帧的图像信息。如图 3.31 所示,计算当前帧(第 i 帧)和前一帧(第 $i-1$ 帧)图像经过分块 DCT 变换后相应直流系数的差值 $D_{DC}(i, j)$ 如下

$$D_{DC}(i, j) = \left| \hat{X}_{i,j,0} - \hat{X}_{i-1,j,0} \right| \quad (3.59)$$

设定阈值 T_F ,若 $D_{DC}(i, j)$ 小于 T_M ,当前帧图像块 c_{ij} 被划分为慢速运动区域;否则, c_{ij} 被划分为快速运动区域,这一类图像块的集合表示为 $Y_i = \{c_{i,f_1}, c_{i,f_2}, \dots, c_{i,f_q}\}$,其中 f_j 表示快速运动块的索引, $1 \leq j \leq q$, q 为快速运动块的总数。对于第 i 帧 c_i 来说,只有那

些包含高细节信息并且沿时间轴变化较快的图像块, 即集合 X_i 和 Y_i 的交集, 被选出来进行水印嵌入, 这里用 $Z_i = \{c_{i,e1}, c_{i,e2}, \dots, c_{i,en}\}$ 来表示这一交集, 其中 e_j 表示高细节快速运动块的索引, $1 \leq j \leq n$, n 为高细节快速运动块的总数。需要注意, 对于视频序列的第一帧, 无法定义运动快慢, 所以只选择那些包含高细节信息的图像块来嵌入水印。阈值 T_D 和 T_F 的选择应充分考虑视频质量和水印鲁棒性两方面因素。

2. 水印生成

嵌入的水印信息通常由单向哈希函数得到, 它可以代表版权所有者的信息或用于盗版跟踪的指纹信息。令水印信息的比特数为 64, 这样一方面可以满足 EBU 的标准, 另一方面与 8×8 的图像块大小一致。水印生成过程如图 3.32 所示。首先, 将原始信息 $m_i \in \{0, 1\}$ 转换成双极性形式 $w_i \in \{-1, 1\}$, 为增加水印处理系统的安全性, 在算法中利用一个由密钥 k 控制的伪随机序列对水印进行置乱。如果无法得到密钥, 即使攻击者知道水印嵌入位置也无法破解水印。接着, 进一步对水印信息进行升维操作, 将一维水印信息重新排列成二维图像块, 这便于以后在嵌入过程中进行位平面替换。这个二维的图案 $w = \{w_{ij}, 1 \leq i \leq 8, 1 \leq j \leq 8\}$ 就是待嵌入水印。需要注意的是, 若将水印信息扩散到多个图像块中, 那么可嵌入更多水印信息位, 同时还可利用扩展频谱和纠错编码技术来提高鲁棒性。

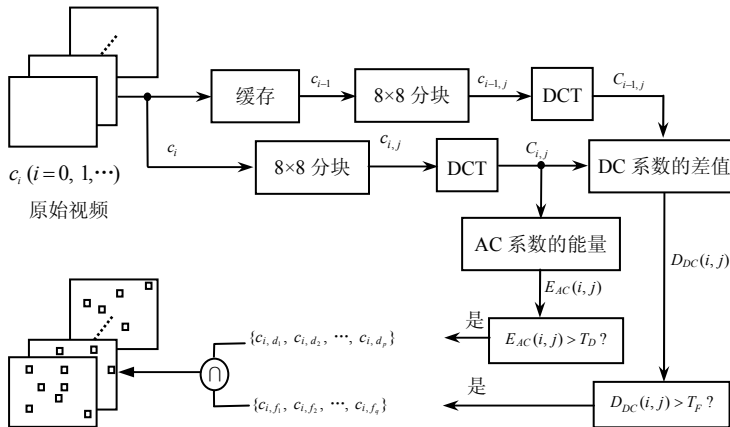


图 3.31 嵌入区域的自适应选择过程

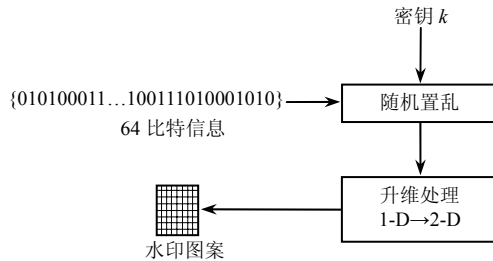


图 3.32 水印生成过程

3. 水印嵌入

这里所给出的水印算法直接在时空域中进行水印嵌入操作, 其基本思想是进行位平面替换。为了克服时空域水印鲁棒性差的缺点, 将水印信息多次嵌入到图像的不同区域。提取时, 根据“多数原则”来恢复水印信息, 从而进一步提高水印的鲁棒性。这里以一帧图像为例, 将水印嵌入到原始视频中的亮度分量。如图 3.33 所示, 首先对选中的

图像块 c_{i,e_l} 进行位分解，得到 8 个位平面（最低位平面被标记为 Number0，最高位平面被标记为 Number7）。然后根据位平面替换对图像质量的影响以及水印的鲁棒性来选择位平面进行水印嵌入。实验结果表明，当水印被放置在第三个位平面或以下时具有很好的不可见性。另一方面，较低的位平面容易受到有损压缩和噪声的攻击，而这些攻击对高位平面的影响相对较小。考虑到所选择的区域具有较好的掩蔽特性，将水印嵌入中间的位平面，这样可保证嵌入的水印同时具有很好的不可见性和鲁棒性。

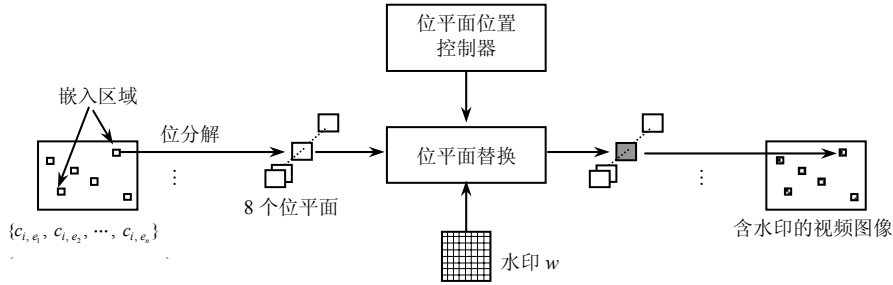


图 3.33 水印嵌入过程

为了进一步提高水印系统的安全性，在嵌入过程中引入一个位平面控制器，即利用伪随机序列发生器的输出随机选择位平面进行替换。这里，候选的位平面主要为第二、第三、第四个位平面。这种控制机制简单并且易于实现，但却可以有效地防止水印被自动去除。在这种情况下，攻击者很难从视频信号中去除水印，除非视频质量被严重破坏。

4. 水印检测及提取

这里给出两种方法从含水印视频中提取水印信息，这两种方法都可以分为两个步骤：水印检测和水印提取。最主要的区别是第一种方法在检测过程中需要原始视频的参与，而第二种方法不需要原始视频。

(1) 利用原始视频作为参考进行水印检测

利用原始视频作为参考，并根据前面所描述的两层检测机制，可找到所有符合条件的水印嵌入区域。通过比较，在含水印视频中找对应嵌入位置 $\{c'_{i,e_1}, c'_{i,e_2}, \dots, c'_{i,e_n}\}$ 。以一帧含水印图像为例，首先将每一个图像块 c'_{i,e_l} , $1 \leq j \leq n$ 按位平面进行分解

$$c'_{i,e_l} = \sum_{l=0}^7 c'_{i,e_l}(l) \times 2^l \quad (3.60)$$

其中， $c'_{i,e_l}(l)$ 代表位平面。分别计算 w 和第二、第三、第四个位平面的相关值

$$r_l = c'_{i,e_l}(l) \otimes w, \quad l = 2, 3, 4 \quad (3.61)$$

这里符号“ \otimes ”表示二维相关运算。在没有受到任何攻击的情况下，相关器的最大响应应该为 64。然后将相关运算后的结果 r_l 与预先设定的阈值 T_c 进行比较， T_c 的选择要同时考虑水印检测的虚警概率和漏检概率。检测结果由下式确定

$$\begin{cases} r_l \geq T_c & \text{检测到水印} \\ r_l < T_c & \text{没有检测到水印} \end{cases} \quad (3.62)$$

若检测器响应大于 T_c ，将相应的位平面保存在集合 $U_i = \{c'_{i,b_1}(l), c'_{i,b_2}(l), \dots, c'_{i,b_u}(l)\}$, $l \in \{2, 3, 4\}$ ，其中 b_j 表示可疑位平面的索引， $1 \leq j \leq u$ ， u 为可疑位平面的总数。该集合用于下一步根据“多数原则”来恢复水印。否则，认为视频数据中没有嵌入水印或水印信息遭到了破坏。

一般而言,有原始视频参与的水印检测算法具有较强的稳健性,因为它可以有效地去除各种噪声信号干扰,并可以抵抗多种形式的攻击,如裁剪、比例缩放、帧去除、帧重组等,从而使得检测结果更加可靠和准确。但从实用角度考虑,原始视频的传输和保存等环节都为水印处理的安全性增加了许多不利因素。对于许多实时应用场合,如数字电视传输和视频点播系统等,存储海量数据代价昂贵并且也是不现实的。下面介绍一种不需要原始视频数据参与的水印检测方案。值得注意的是,这里并不需要知道有关时间轴的一些信息,也就是说无需知道视频序列的起始位置和顺序,可以实现随机检测。

(2) 不需要原始视频参与水印检测

与第一种检测方案相比,第二种检测方案稍有不同。针对含水印的视频图像,在帧内利用下式检测图像块中是否包含细节信息

$$E'_{AC}(i, j) = \sum_{l=1}^{63} (C'_{i,j,l})^2 \quad (3.63)$$

将计算得到的结果与阈值 T'_D 进行比较,由此得到候选集 $V_i = \{c'_{i,h_1}, c'_{i,h_2}, \dots, c'_{i,h_v}\}$, 其中 h_j 表示候选块的索引, $1 \leq j \leq v$, v 为候选块的总数。需要注意的是,这里 $T'_D < T_D$ 。然后,对于集合 V_i 中的每一个图像块先进行位平面分解,接着进行相关检测。如果相关运算的结果大于等于 T_c ,则表示检测到水印存在。类似地,将相应的位平面保存在集合 U_i 中,用于第二步水印提取。

在许多情况下,攻击者可能通过插入或删除一帧图像来改变视频序列的长度,或者通过帧重组来改变视频序列的顺序,这些都对不需要原始视频参与的水印检测算法提出更高的要求。嵌入一个同步信号将有助于检测,但是这同样增加了受到攻击的风险。这里所给出的细节信息检测机制依赖于图像的内容,而不是视频信号的三维结构,所以该算法可以抵抗沿时间轴破坏同步关系这一类攻击手段。

(3) 水印提取

当水印嵌入位置确定后,可以直接进行水印提取。考虑到嵌入水印的冗余性以及嵌入位置的多样性,水印提取算法采用加权综合的方法,并利用“多数”原则来恢复水印信息。水印提取过程可以用下式表示

$$w' = \sum_{U_i} \frac{g_{i,e_j}(l)}{\sum_j g_{i,e_j}(l)} c'_{i,e_j}(l) \quad (3.64)$$

其中 $g_{i,e_j}(l)$ 表示每一个位平面 $c'_{i,e_j}(l)$ 的权值,它的定义式为

$$g_{i,e_j}(l) = 2^l, \quad l \in \{2, 3, 4\} \quad (3.65)$$

同样,也可以从一段视频中(以任意位置为起始点)提取水印

$$w'' = \sum_i \sum_{U_i} \frac{g_{i,e_j}(l)}{\sum_i \sum_j g_{i,e_j}(l)} c'_{i,e_j}(l) \quad (3.66)$$

将加权综合后的结果转换为二值信息,并利用嵌入过程中所使用的密钥 k 进行伪随机置乱逆操作,从而由 w' 或 w'' 恢复出信息 m' 。

5. 仿真实验

实验中采用 MPEG-2 标准测试序列“Mobile & Calendar”、“Table Tennis”和“Coast Guard”进行水印嵌入和提取。视频序列为 CIF (Common Intermediate Format) 格式,每一帧大小为 352×288 , 长度为 30 帧。图像由亮度分量 Y 和两个色差分量 Cb、Cr 构成,

色度格式 (Chroma Format) 为 4:2:0。在实验中仅考虑视频序列的亮度分量。

图 3.34 (a) 为原始视频序列中连续的三帧图像, 图 3.34 (b) 为含水印视频序列中相应的连续三帧图像, 图 3.34 (c) 显示的是对应图像的差值 (为了便于显示, 将差值进行了放大, 较亮的区域对应较大的差值)。表 3.7 列出了含水印的“Table Tennis”序列各帧图像的 PSNR 值。可以看出, 嵌入的水印具有很好的不可见性, 含水印视频中各帧图像 PSNR 的最大值、最小值和平均值分别为 44.4dB、42.6dB 和 43.5dB。

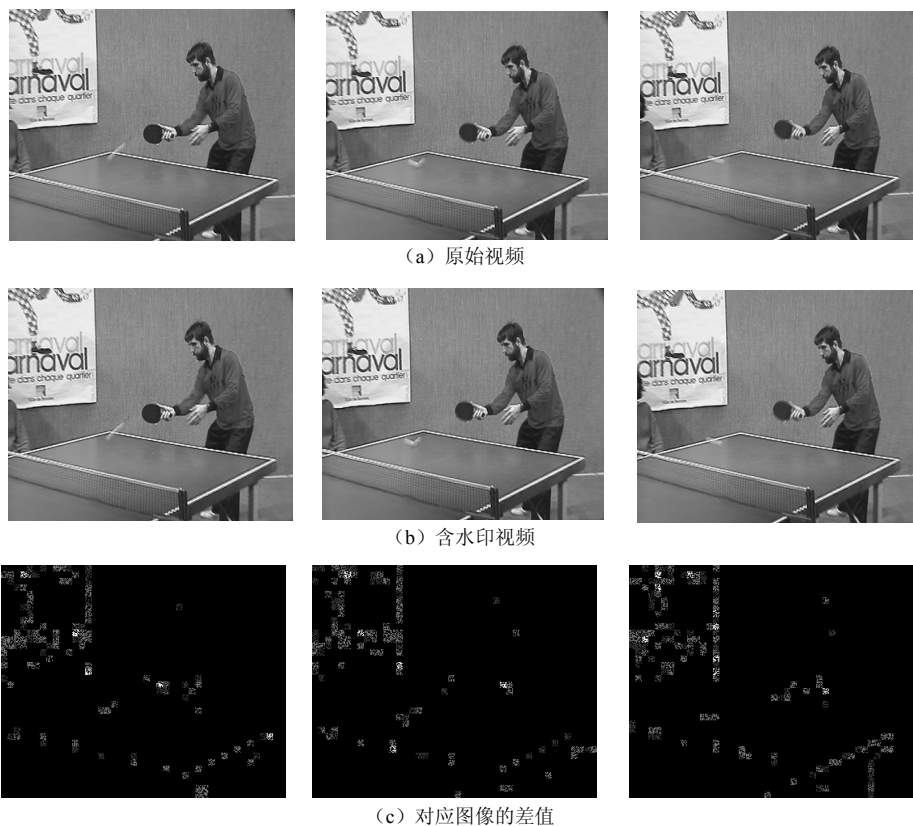


图 3.34 “Table Tennis” 视频片段中的连续三帧

表 3.7 含水印的“Table Tennis”序列各帧图像的 PSNR 值

帧 数	PSNR (dB)									
第 0~9 帧	43.23	44.07	44.39	43.63	43.24	43.66	43.52	43.55	43.22	42.90
第 10~19 帧	42.63	43.74	43.17	43.95	43.03	42.57	42.86	43.97	43.63	43.57
第 20~29 帧	43.88	43.10	44.32	43.30	43.07	43.57	43.63	43.64	44.03	43.14
平均值	43.47									

由于视频通常以压缩格式进行存储和传输, 因此需要验证水印对于 MPEG-2 有损压缩的鲁棒性。这里采用 MPEG 工作组提供的标准编解码器来实现 MPEG-2 压缩编码以及相应的解码过程, 表 3.8 给出了经过 MPEG-2 攻击后的水印提取结果。由实验结果可以看出经过 MPEG-2 压缩后仍可从含水印的视频图像中正确恢复出水印信息。当编码码率下降到 4Mbit/s, 对应压缩比为 18.3 时, 从单帧图像中提取水印有少量的水印信息无法正确恢复。该算法可以进一步通过采用抗干扰编码技术来提高水印的鲁棒性。

表 3.8 经过 MPEG-2 攻击后的水印提取结果

算 法	测试 序列	MPEG-2 编码码率	提取水印信息的正确率			
			从图像组中提取	从单帧图像中提取		
				最小值	最大值	平均值
需要原始视 频参与的检测 和提取算法	Mobile & Calendar	6M bits/s	100%	100%	100%	100%
		4M bits/s	100%	96.9%	100%	98.8%
	Table Tennis	6M bits/s	100%	100%	100%	100%
		4M bits/s	100%	98.4%	100%	99.5%
	Coast Guard	6M bits/s	100%	100%	100%	100%
		4M bits/s	100%	95.3%	100%	98.6%
不需要原始 视频参与的检 测和提取算法	Mobile & Calendar	6M bits/s	100%	100%	100%	100%
		4M bits/s	100%	95.3%	100%	98.1%
	Table Tennis	6M bits/s	100%	100%	100%	100%
		4M bits/s	100%	96.9%	100%	99.2%
	Coast Guard	6M bits/s	100%	100%	100%	100%
		4M bits/s	100%	93.8%	100%	98.3%

在实际应用环境中，视频信号还有可能经受其他形式的攻击。例如，在录制视频信号中，由于一些异常操作常常会引入噪声或出现丢帧情况，恶意的攻击还包括插入某一帧、帧重组等形式。仿真实验结果表明，本小节介绍的自适应水印处理算法对于加性高斯噪声、丢帧、插入某一帧、帧重组等攻击形式具有很好的鲁棒性。另外，由于水印嵌入在亮度分量中，所以当视频信号从 YCbCr 空间转换到 RGB 空间、从彩色图像转换到灰度图像时，对水印的提取结果没有影响。

经过分析发现，水印提取过程需要原始水印信息的参与，并不是真正意义上的盲检测。为此，可以在改进算法中利用直接序列扩频技术来生成水印，使得水印提取过程完全由密钥控制。利用伪随机序列（ m 序列或 Gold 序列）作为扩频码，采用类似于二相相移键控（Binary Phase Shift Keying, BPSK）的调制方法，对水印信息 w'_i 直接进行调制。具体而言，就是将每一位水印信息重复 64 次，然后直接与扩频码相乘，从而实现水印信息的频谱扩展。采用这种方法可将一位水印信息分布于一个 8×8 的图像块中，并根据所选择图像块的多少确定嵌入的信息量。这里将扩频信号直接加入到被选中的图像块 $c_{i,el}$ 中，信号增益 G 的取值范围为 $4 \sim 8$ 。这样，在取得较好视觉效果的情况下又保证了一定的鲁棒性。水印检测过程中，首先对含水印图像进行去相关滤波，然后再进行相关检测。水印检测和提取算法的安全性完全依赖于密钥的选择，也称为真正意义上的盲提取水印方案（Oblivious Watermarking）。这里选取“Table Tennis”和“Coast Guard”两组测试序列进行水印嵌入和提取实验，并根据 MPEG-2 编码中不同的图像类型对结果进行统计平均。表 3.9 给出了含水印视频经过 MPEG-2 有损压缩后的水印提取结果。由仿真结果可以看到，改进算法对于 MPEG-2 攻击同样具有较好的鲁棒性，其中从 I 帧中提取水印的正确率最高，P 帧次之，B 帧最低，这与相应类型的图像的编码质量有关。通过仿真实验发现，预测滤波器的选择对于水印检测结果有很大影响，因此今后将针对如何设计自适应的预测滤波器这一方面进行研究，以进一步提高水印检测的性能。

表 3.9 改进方法经过 MPEG-2 攻击后的水印提取结果

MPEG-2 编码码率	提取水印信息的正确率					
	Table Tennis			Coast Guard		
	I 帧	P 帧	B 帧	I 帧	P 帧	B 帧
6M bps	96.5%	94.1%	92.6%	99.8%	98.6%	98.1%
5M bps	93.8%	92.4%	89.5%	98.3%	97.2%	95.9%
4M bps	89.2%	86.3%	83.3%	94.2%	93.2%	89.1%

3.7.4 典型压缩域视频水印算法

在视频中嵌入水印一般考虑 MPEG 编码标准。在 MPEG 编码标准中，有 3 种图像类型：内部编码帧（I 帧）、前向预测帧（P 帧）和双向预测帧（B 帧）。I 帧的编码类似于 JPEG，利用帧内相邻像素间的空间冗余来压缩信息；P 帧编码时要用到先前的帧，当前的帧又可作为后面预测帧的参考帧；B 帧的数据压缩效果最显著，它的预测需要先前帧和后续帧的信息，且自身不能作为其他帧的预测参考帧。P 帧和 B 帧都利用了相邻帧间的时域冗余来压缩信息，同时预测误差信号还可进一步的去除空间域冗余。去除空间域冗余主要用到了 DCT、量化和熵编码技术；去时域冗余用到了运动补偿、运动表示和运动估计等技术。在压缩视频流中嵌入水印的算法，可以按嵌入位置划分为在离散余弦变换系数中嵌入水印、在运动矢量中嵌入水印、在 MPEG-4 脸部运动参数中嵌入水印和在 VLC 域嵌入水印。下面对各种压缩域视频水印嵌入方法予以简要介绍。

1. 在 DCT 系数嵌入水印

在 DCT 系数中嵌入水印是目前研究最多的算法，可以借鉴图像 DCT 水印的成果，这方面的算法也很成熟。下面简要介绍两种典型的方法。

(1) 扩频方法

Hartung 和 Girod^[43]提出利用扩频思想在 MPEG-2 压缩视频中嵌入水印的算法。水印信号经过扩展、放大和调制，得到一个伪随机序列；然后对其进行 8×8 的 DCT，并将 DCT 系数叠加到 MPEG-2 码流的 8×8 的 DCT 系数上。它主要考虑两个问题，① 由于 MPEG-2 的 DCT 系数是用变长编码进行编码的，系数在添加水印前后的编码长度会发生变化。因此，如果要求不增加视频码流长度，那么，在出现添加水印后 DCT 系数的编码比特数增加的情况时，则仍将保留原有的系数。② 在 MPEG-2 编码方式中，帧间编码帧（P 帧和 B 帧）是从其他帧预测得到的，用一个运动补偿向量来从其他帧重建当前帧，P 帧本身也可能作为其他帧的预测参考。一个帧内的微小变化会在时间、空间上传播开来。因此，在水印信号之外，需添加一个偏移补偿信号来补偿前一帧的水印信号。

(2) 自适应方法

Simitopoulos 等^[44]提出了一种在 MPEG 流压缩域嵌入水印的算法，把视觉分析和块分类技术结合起来，自适应选择 I 帧亮度模块 DCT 域的量化 AC 系数嵌入水印。水印系数 $w(i, j)$ 是伪随机序列 (± 1)。 $w(i, j)$ 和对应的量化嵌入标志 $q(i, j)$ 、分类标志 $t(i, j)$ （分别由视觉分析和块分类过程产生）的乘积加到每个被选量化参数上。加入水印后的系数为

$$S(i, j) = C(i, j) + w(i, j) \cdot q(i, j) \cdot t(i, j) \quad (3.67)$$

其中， $C(i, j)$ 为原始 DCT 量化系数， $S(i, j)$ 为含水印 DCT 量化系数。该算法的优点是，在对 DCT 系数量化之后嵌入水印，由于量化之后执行 MPEG 编码是无损操作，因

此任何嵌入的信息不会在后续处理过程中丢失。这样,当执行检测过程时,水印信息就会完整地存在于量化系数中。

2. 在运动向量中嵌入水印

Kutter 等^[45]在一份 MPEG-4 提案中提出了一种直接针对 MPEG-4 编码视频流的水印算法,并通过修改运动向量来嵌入信息。水印嵌入过程如下,在一个运动向量的某个分量,比如垂直分量 v 中嵌入水印,设 $b=\{0,1\}$ 为待嵌入的比特值,水印嵌入规则为

$$v' = \begin{cases} v + \delta & \text{若 } (v \times q + T) \bmod 2 \neq b \\ v & \text{其他} \end{cases} \quad (3.68)$$

其中, $T=2 \times$ 运动估计搜索窗口, $\delta=(2n+1)/q$, n 为整数。一般地,对于空间运动向量, $n=1$; 否则, $n=0$ 。 q 指定了对运动向量修改的范围。对应的水印提取规则为

$$b = (v \times q + T) \bmod 2 \quad (3.69)$$

通过实验, q 可选 1 或 2, 取 1 时对于压缩鲁棒性较好。每帧随机选取一块,在每个运动向量可嵌入 2 比特。计算复杂度几乎可以忽略,对于帧的比特率的影响也是非常小的。

为了改进上述方案,可以只在运动矢量的幅值较大的宏块嵌入水印,并且水印嵌入在运动矢量相角变化较小的分量上,具体算法如下,① 从视频流得到运动矢量;② 计算运动矢量的幅值;③ 选择要嵌入水印的宏块,并计算该宏块的运动矢量的相角 θ ;④ 根据 θ 的值确定在水平分量还是垂直分量嵌入水印。若 $\theta < 45^\circ$,则在水平分量嵌入水印;若 $\theta > 45^\circ$,则在垂直分量嵌入水印;若 $\theta = 45^\circ$,水平和垂直分量都嵌入水印。

3. 在脸部运动参数中嵌入水印

Hartung 等^[46]提出了在 MPEG-4 脸部运动参数中嵌入水印的算法,仍然采用了扩频通信的思想。在 MPEG-4 中定义了一个一般的脸部,并能够通过**脸部运动参数**(Facial Animation Parameter, FAP)运动起来。FAP 共有 66 个,包括整个头部的运动参数(头的倾斜和偏转角度)和局部的脸部运动参数(眼睑的张开、嘴唇的张开、嘴角的运动)。在 MPEG-4 编码过程中从视频序列中确定 FAP,可以将 FAP 看成随着时间变化的 d_{\max} 维向量 $FAP(t)$ 。 d_{\max} 是所传输 FAP 的数目 ($d_{\max} \leq 66$), t 是视频帧的整数时间索引值, $FAP_i(t)$ 是在时间 t 第 i 个 FAP ($i \in \{1, 2, \dots, d_{\max}\}$)。水印算法的基本思想是将 1 比特水印信息散布到多于一个 FAP 中,如 $A \times B$ 的 FAP 块中。对要嵌入的比特信息,进行扩展、调制;然后进行低通滤波和振幅调制,最后加到所选择的 FAP 块中。

从含水印的 FAP 中提取嵌入的水印,需要从含水印的 FAP 中减去原始的 FAP;再用和水印嵌入相同的伪随机序列进行相关运算,然后判断是否大于给定的阈值。但在某些情况下,不能直接得到 FAP,而只能得到播放的视频序列;这时需要首先从播放的序列中估计出 FAP,然后从估计的 FAP 中提取水印。

该算法存在的主要问题包括:需要原始载体视频,且水印提取出来的速率不是均衡的;在水印嵌入和提取的处理中如何考虑人类视觉系统的特性。

4. 在 VLC 域嵌入水印

由于视频的数据量非常大,在原始视频中嵌入和提取水印不容易满足实时性的要求。Lu 等^[47]提出的针对 MPEG-2 视频流的水印算法,它是在**可变长编码**(Variable Length Coding, VLC)域嵌入水印,并实现实时检测。相对于解码过程所需时间,水印检测所需的时间可以忽略不计。该算法是基于携带辅助信息通信的思想设计的。

在视频解码阶段,通过 VLC 解码,视频流被解码成码字,每个码字表示为 (r, l) , 其

中 l 值是经量化后不为零的 DCT 系数, r 值表示该 l 值前面连续零值的个数。水印只嵌入在 l 值上, 不改变 r 值, 这是由于改变 l 值, 只影响一个 DCT 系数; 而如果改变 r 值, 就会有很多 DCT 系数被迫改变, 图像的失真将非常严重。考虑到压缩流中 GOP 的结构由 I 帧、B 帧和 P 帧组成, B 帧和 P 帧的重构依赖于 I 帧, 因此只在 I 帧中嵌入水印就足够了, 解码后非 I 帧就自然包含水印。由于在不同的压缩率下(r, l)对的个数会改变, 从而引起水印检测不同步问题, 因此选择在宏块中隐藏水印, 对不同的攻击(包括压缩)宏块的数目是不变的。水印嵌入在 I 帧的 Y 分量中。

该算法对不同位率的 MPEG 压缩、添加加性噪声、锐化、帧平均、帧率的改变、均值滤波、I 帧删除+压缩复合攻击、共谋攻击等攻击具有鲁棒性, 但对复制攻击的鲁棒性差。另外, 如果嵌入过程在一些小的 DCT 系数上进行, 含水印视频将不具有鲁棒性。

3.8 本章小结

本章主要讲述了信息隐藏领域的第二个主要研究分支——数字水印技术。首先介绍数字水印技术的提出背景, 然后介绍数字水印技术的相关概念、与隐写术的区别和数字水印技术的分类。接着, 介绍数字水印系统的基本框架、通信模型和几何模型, 然后介绍数字水印技术的应用、特性和评价问题。最后, 介绍基于图像、音频和视频这三种主要载体的数字水印技术, 对于每一种载体, 都从其水印系统的基本要求入手, 先介绍其水印系统模型和算法分类, 然后介绍评价指标, 最后分类介绍典型的水印算法。针对图像载体, 分鲁棒算法和脆弱算法进行介绍; 对于音频载体, 分时域算法、变换域算法和压缩域算法进行介绍; 对于视频载体, 分原始域算法和压缩域算法进行介绍。

实际上, 除了图像、音频和视频载体外, 还有许多载体作品需要版权保护和内容认证, 如动画、三维模型、文本、关系数据库、软件、网页等。下面简要列举三种典型载体水印技术。

(1) 目前大量的三维数字产品模型已经广泛而深入地应用到产品研发的协同设计、虚拟维修等过程中, 所涉及的模型既包括与自身密切相关的模型, 如几何模型等, 也包括与虚拟应用环境相关的模型, 如活动模型等。如何在既充分利用网络的优势实现便捷的信息共享的同时, 又保证三维数字模型信息的安全性、完整性已成为亟待解决的问题。三维模型数字水印技术为解决此类问题提供了一种有效途径。三维模型水印的挑战性在于: ① 顶点的不规则性和无序性增加了水印嵌入的难度, 三维模型的顶点没有自然的排列顺序, 所以大部分三维模型水印算法不得不借助原始模型, 通过网格对齐和重采样进行非盲检测, 或者只考虑顶点排列顺序不变的情况; 而且三维模型顶点的分布是不规则的, 缺少频谱分析所需要的自然参数化方法。② 复杂多样的攻击手段使三维模型水印的检测异常困难。平移、旋转、均匀缩放、重排序等操作对三维模型属性没有进行任何改变, 却严重影响了水印的检测。噪声、滤波、重新三角化、重采样、剪切等进一步增加了三维模型水印检测的难点。

(2) 相比于多媒体载体, 自然语言文本具有冗余度低、语言规则复杂、计算机处理困难等特点。关于文本的信息隐藏已经在上一章中介绍过, 这里稍微再强调一下。文本水印呈现起步较晚、实用成果较少的现状。早期的文本水印带有图像的影子, 通过微小地改变文本的视觉特性嵌入信息, 比如字移、行移、笔画粗细修改、形态修改等。随后, 为抵御重排版攻击, 人们又发展出自然语言水印, 根据对载体的改动方式又可分进

一步划分为基于语法的、基于语义的和生成文本的三类。与排版类水印相比,自然语言水印改变的是文本的内容,只要内容不变水印就能保留,具有良好的鲁棒性,同时保持文本的风格、含义、感情色彩等基本不变,对载体的影响降低到最小程度,也增加了检测的难度。相对于自然语言水印的嵌入算法,针对自然语言水印的攻击成果更少。攻击包括对水印的检测、对水印的破坏和对水印的还原。这方面的研究有待进一步深入和发展。

(3) 随着关系数据库的广泛使用,随之产生了在关系数据库中嵌入水印信息的需求。如对那些提供信息服务(包括气象信息、医疗信息、人才市场信息、股票交易信息电子元器件参数信息等)的机构,其主要资产便是存储于数据库里的大量数据。通过在关系数据库中嵌入代表所有权的水印信息,可以将数据库与其拥有者联系起来,从而实现数据库的版权保护。由于数据库数据量大但冗余小,需要对数据进行一定的正常维护操作等特点,不能将传统多媒体数字水印技术直接应用到数据库中。必须研究数据库和数字水印技术的特点,开发一种满足关系数据库版权保护的水印技术,来解决数据库版权保护问题。与多媒体数据相比,关系数据库数据有其特殊性。基于这些特殊性,关系数据库数字水印技术有很大的难度。目前尚有以下问题需要解决:大多数数据库水印采用的是空域方法,如何找到一种类似多媒体技术的频域技术来提高水印的鲁棒性;关系数据库中的数据大体上可分为数值型和非数值型数据。目前大多数针对数值类型的数据,如何对非数值型数据进行数字水印操作,进一步扩展数字水印的应用。



习题

1. 请阐述数字水印技术和隐写术的区别和联系。
2. 请阐述数字水印系统的各种通信模型,并进行比较。
3. 请阐述数字水印系统几何模型的各个空间或区域的概念。
4. 请阐述数字水印系统中各种水印生成机制。
5. 混沌水印的生成通常可采用一个简单的混沌系统 Logistic 映射来实现,它可定义为

$$x_{i+1} = \mu x_i (1 - x_i) \quad \mu \in [1, 4], i = 0, 1, 2, \dots \quad (3.70)$$

初始值选为 $0 < x_0 < 1$, 这样一来得到的序列 x 的取值范围是单极性的, 且 $0 < x_i < 1$ 。取初值为 $x_0 = 0.25$, 采用 $\mu = 3.93$ 。为了将生成的实数序列 x 转化为二值序列 p , 可以 0.5 为阈值, 若 $x_i \geq 0.5$ 则对应的 $p_i = 1$; 否则 $p_i = 0$ 。试用 Matlab 或 C 语言编程序, 打开一幅 64×64 的二值图像 m , 生成长度为 4096 的二值混沌序列 p , 由此生成水印 $w = m \oplus p$, 其中 \oplus 表示异或操作, 观察所生成水印信息的伪随机特性。

6. 请阐述数字水印系统中各种水印嵌入机制。

7. 试用 Matlab 或 C 语言编程序, 打开大小为 256×256 的 256 灰度 Lena 图像作为载体图像, 对载体图像进行 8×8 分块 DCT 变换; 然后打开一幅 32×32 的二值图像作为原始信息, 将此原始信息进行伪随机置乱生成水印; 在每块 DCT 系数块中嵌入 1 比特水印信息, 采用替换规则: 选取一对相邻中频系数 c_i 和 c_j 。若 $c_i < c_j$ 且待嵌入水印位为 1, 则把 c_i 和 c_j 交换位置; 如果 $c_i \geq c_j$ 且待嵌入水印位为 0, 则交换 c_i 和 c_j 位置。其他情况则保持 c_i 和 c_j 不变。试着对含水印图像进行 JPEG 压缩、模糊、中值滤波、亮度和对比度变换等攻击, 计算提取的水印和原始水印之间 NC 值, 由此说明算法的鲁棒性。

8. 请解释音频信号的时域掩蔽效应，在音频水印中通常采用哪一种效应？
9. 试述利用三个相邻音频数据块能量关系来嵌入水印的方法，并用 16bit、单声道、采样频率为 22kHz 的“Windows XP 启动.wav”音频文件作为载体对象在 Matlab 或 C++ 中实现该算法。
10. 试比较前置式、内置式和后置式视频水印方法的优缺点，并画出示意图。
11. 在图 3.31 的原始域视频水印算法中，每帧图像的嵌入位置的选取可以基于图像块 DCT 变换系数进行两种准则的判别：运动检测准则与细节检测准则。请解释这两个准则的内涵。

第4章

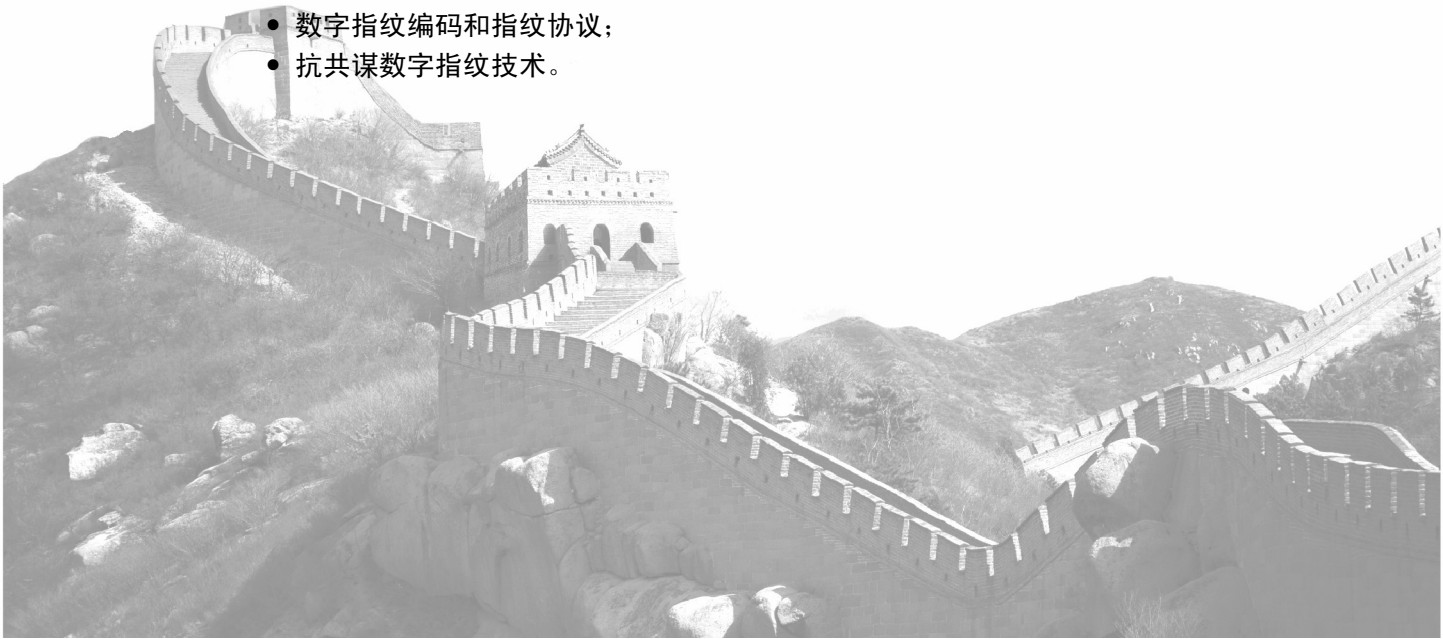
数字指纹技术

本章引言

随着多媒体和通信技术的快速发展，越来越多的多媒体数据如图像、音频、视频等通过网络广泛传播，这就迫切需要确保数据内容的合理分发和正当使用，其技术支持主要涉及数字水印和数字指纹技术。在这些技术中，拷贝检测是基于标识嵌入技术的，即出版商在出售数字产品之前，将不可见的标识嵌入到数据中。数字水印技术将相同的标识嵌入到同一个数字产品中，而数字指纹技术（Digital Fingerprinting）将不同的标识嵌入到同一个数字产品中去，以区分不同的用户。因此数字指纹特别适用于追踪非法散布数据的授权用户。本章从数字指纹技术的提出背景入手，首先介绍数字指纹技术的相关概念、分类、框架模型和性能评价，然后概述最重要的指纹编码和指纹协议问题，接着详细介绍基于不同指纹协议的数字指纹技术，最后讨论抗共谋攻击的指纹编码问题。

本章重点

- 数字指纹技术的相关概念和分类；
- 数字指纹系统模型和性能评价；
- 数字指纹编码和指纹协议；
- 抗共谋数字指纹技术。



4.1 数字指纹技术的提出背景

4.1.1 指纹和指纹识别

数字指纹和数字指纹技术的概念借鉴自传统的指纹和指纹识别。因此，有必要简要回顾指纹和指纹识别的概念。**指纹**（Fingerprint）是灵长类动物和人类手指末端指腹上由凹凸的皮肤所形成的纹路，或是纹路在物体上印下的印痕。这些印痕最常在犯罪学、法医学上被当作证据。指纹重复的机会极微，目前尚未发现有不同的人拥有相同的指纹。广义的指纹包括手掌纹、脚纹、脚掌纹。亨利氏指纹分类法依照指纹形状将指纹分成斗形纹、箕形纹和弧形纹三大类，如图 4.1 所示。

指纹的利用与发展已有几百年历史，能让读者陶醉的也许只有犯罪现场的指纹鉴定。古代人对研究手指指纹非常感兴趣，在很多地方如陶器和雕像上都可发现指纹踪迹。研究范围相当广泛，如发现的新石器时代的花瓶，青铜时代的锅，亚述人泥片，墨西哥陶器和阿芝台克黏土图片。很显然，这些事例中的指纹是在当时器具的制造过程中留下的。**指纹识别**（Fingerprinting）作为专业术语出现在很早以前，但很难追溯历史上是谁首先使用的。关于指纹比较确切的历史记录如下：1823 年，德国学者 Johannes E. Purkinje 将指纹依不同纹形加以分类，有了斗形、箕形、弧形的分类法，后人据此对指纹进行了更详细准确的定义；1880 年，Henry Faulds 博士在 Nature 国际期刊上发表第一篇有关指纹的研究，1886 年，他将这个想法提供给伦敦的警察单位，但不被采用；1891 年，阿根廷警官 Juan Vucetich 在阿根廷首创世界第一个罪犯指纹档案；1892 年，Fancis Galton 在他的新书《指纹》中对发表了对指纹更为详细的分析、鉴定和法医学上的运用；1901 年，Edward Richard Henry 设计了一套指纹识别系统，并在英格兰和威尔士率先使用；1902 年，在纽约，Henry P. DeForrest 博士将指纹识别系统运用在市民服务上。

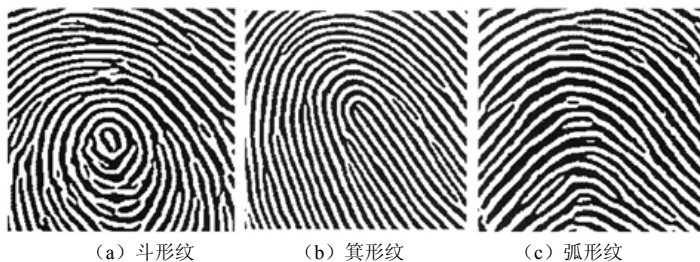


图 4.1 指纹形状

4.1.2 数字指纹技术的提出背景

据国际电信联盟（ITU）报告，目前全球互联网网民数量已经突破 27 亿。据国内报告，截至 2013 年 12 月底，中国网民数量已突破 6 亿。互联网的迅速普及，使信息的存储和传播方式发生了巨大的改变，以数字形式存在的各种作品及其信息网络的传播方式，给版权相关产业带来了巨大的发展机遇，同时也使互联网版权及相关管理工作面临着历史性的挑战，使作品版权的私人专有性和集体共享性之间的利益博弈表现得更加突出和尖锐。互联网的虚拟性和自由性决定了侵权的低成本与无节制。网络传播信息在技术上几乎无法限制，越来越多的多媒体数据如图像、音频、视频等通过网络广泛传播，

作品可以被很方便、精确、逼真地进行复制,或被任意删改或者移植。这种盗版行为不仅使作者利益受到严重侵害,对提升全社会的创造、创新、创作能力产生负面影响,也制约着网络环境下版权产业的健康发展。因此,迫切需要采取技术手段确保数据内容的合理分发和正当使用。以 DRM (Digital Rights Management) 为主线的数字版权管理技术以及以避风港原则为代表的法律制度框架无法根本解决互联网版权保护问题。数字版权产业亟待模式创新,所谓的模式创新不是单一的技术手段或者管理手段的革新,而是集成运用技术创新成果和标准的引领作用进行体系化的互联网版权综合治理。法国、韩国等国家依据互联网时代数字版权保护的具体特点,纷纷提出了不同的创造性管理模式以及配套的专门机构和法律制度,通过法律创新、模式创新、机构创新和技术创新,有效应对互联网版权保护面临的挑战。

在上述背景下,以标识技术为核心的数字版权技术创新成果被广泛采用。数字指纹技术和数字水印技术是近几年发展起来的新型数字版权标识技术。数字水印技术是向数字产品中嵌入版权拥有者的信息,当发生争议时能够确认版权的归属,同一作品嵌入的水印信息是相同的。而数字指纹技术是在产品中嵌入数字水印的基础上再嵌入的用户个人信息,产品提供者能够根据该信息对非法用户进行追踪。对于用户购买的同一功能的数字产品,嵌入的信息是不同的。借用指纹识别的概念,数字指纹技术的概念最早是由 Wagner 在文献[48]中把它作为一种在多媒体电子数据中保护版权者权利的方法提出来的。众所周知,指纹是指一个产品所具有的模式,它能把一个产品和同其他同类相似产品相区分。同样,人们希望能用数字指纹来区分数字产品的不同复制。当然,与数字指纹技术相关的领域还包括隐写术 (Steganography) 和叛逆者追踪 (Traitor Tracing) 技术。

数字指纹技术具有广泛的应用环境和广阔的应用前景。它可以用于在线出版业,如电子图书馆的构建;随着数字电视和数字广播的发展,它可应用于 DVB (Digital Video Broadcast)、VOD (Video on Demand) 等环境下付费数据的保护。数字指纹技术与数字水印技术、数字签名和数据加密等技术一样是重要的内容保护技术。例如,法国 HADOP 机构采用创新的指纹技术,由版权人提供作品唯一指纹,形成作品指纹数据库,并要求网络服务商安装指纹识别系统,通过指纹锁定作品的权利状态、使用情况和非法上传/下载行为,从而建立起以事前利益分享机制为核心的新模式和扼制网上侵权盗版行为的有效防线。

作为一种新兴的版权保护技术,数字指纹技术的研究目前已经取得一定的进展,并且在国内外得到一定关注。在数字指纹技术的研究方面,国外开始得比较早,且已经取得一些成果。一些从事信息隐藏技术研究工作的公司或团体纷纷推出相关的软件,日本电气公司、日立制作所、先锋、索尼及 IBM 等公司正在努力开发具有统一标准的数字指纹技术。近期研究最活跃的是美国马里兰大学 (Maryland University) 研究团队,他们对包括数字指纹技术在内的数字多媒体信息安全技术进行了一系列深入研究。我国在数字水印技术方面的研究近些年比较活跃,然而数字指纹技术的研究与国外还有很大差距,因此进一步加强我国的版权保护技术研究的力度和深度势在必行。

4.2 数字指纹技术的相关概念和分类

数字指纹技术和数字水印技术是近年来出现的新型数字版权保护手段,可以统称为版权标识技术。数字水印技术是向数字内容中嵌入代表版权拥有者身份的相关信息,确定自己的版权归属权。而数字指纹技术是在数字内容中嵌入与用户有关的信息,分别向

分发的不同用户复制中嵌入不同的信息，嵌入的指纹对不同购买者是不同的。此时版权拥有者能够根据信息追踪非法用户，当发生非法复制时，版权拥有者能够确定侵权用户。简单来说，数字水印技术通常用来确定原始作者，数字指纹技术通常用来进行复制追踪。事实上，数字指纹技术又是广义数字水印技术的一种具体应用。

4.2.1 数字指纹技术的相关概念

1. 数字指纹和数字指纹技术的术语和定义

指纹技术（Fingerprinting）是一种很古老的技术，早在几百年前，人们就用它来保护对数表。其基本思想是：对于一个随机数 c ，发行商在其对数值 $\log c$ 的某一个不太重要的数位中，比如小数点右边第 n 个位置，引入一个很微小的误差。这样一来，一旦有购买者非法出售其所购买的对数表，发行商就可以通过找出非法出售的对数表中那些微小的误差，来追查原始的购买者。在现代社会，物理指纹技术已经得到了广泛应用，它已经用于安全身份认证、公司员工考勤以及刑事侦察等多个领域。而在版权保护领域，为避免未经授权的数字作品复制制作和发行，出品人将标识不同用户的符号序列作为指纹嵌入作品的合法复制中。一旦发现未经授权的复制，就可以根据此复制所恢复出的指纹来确定它的来源，即盗版者（叛逆者）。这就是**数字指纹技术**（Digial Fingerprinting），而把嵌入的信息称为**数字指纹**（Digital Fingerprint）。数字指纹技术把具体的数字内容与用户身份相绑定，是帮助我们追踪盗版的关键技术。

1983 年，Wagner 发表了题为“Fingerprinting”的文章，比较全面地介绍了指纹技术的基本思想和一些相关术语，同时给出了一些使用指纹技术的实际例子^[48]。Wagner 认为指纹在现实世界中是普遍存在的，任何可能被非法使用的对象都可以给它添加一个指纹，使得在它被非法使用（盗版）后，能够找到该对象的原始盗版者。文献[49]扩展了指纹的概念，对指纹在数据保护中的应用做了进一步的研究。数字指纹技术是在指纹技术的基础之上发展起来的一种新型数字版权保护技术，并广泛用于不同场合的版权保护，在题为“Collusion Secure Fingerprinting for Digital Data”的经典文献中，作者系统介绍了有关数字指纹技术的一些基本概念与术语，奠定了数字指纹技术的研究基础，被称之为数字指纹技术的开山之作^[50]。与数字指纹技术相关的一些术语如下。

（1）标识

标识是产品的一部分并有若干个可能的状态。

（2）指纹

指纹是标识的集合。

（3）发行商

发行商是一个授权提供者，他将嵌入指纹的产品提供给用户。

（4）授权用户（合法用户）

授权用户是一个获得使用某一嵌入指纹产品的个人。

（5）攻击者（非法用户）

非法用户是非法使用嵌入指纹产品的个人。

（6）叛逆者（盗版者）

叛逆者是非法发行嵌入指纹产品的授权用户。

2. 数字指纹技术与数字水印技术的区别和联系

版权标记技术根据标记内容和所采用技术的不同可以分为数字水印技术与数字指纹

技术，它们之间既有联系又有区别。

首先，它们的侧重点不同。传统的数字水印技术是把相同的、标记版权的特殊信息（数字水印）嵌入到所有的数字产品中，相同的数字产品拥有相同的水印信息。当发生版权纠纷时，可以通过提取预先嵌入的水印信息，能够有效地确认版权归属。因此，数字水印技术主要用于版权证明。数字指纹技术是把不同的指纹信息（序列码）嵌入到不同消费者所购买的数字产品中，不同消费者所得到是内容相同、但嵌入了不同指纹信息的指纹复制。当发生盗版时，内容提供商能够提取预先嵌入的指纹信息，找到相应的盗版者。因此，数字指纹技术主要用于盗版者追踪（叛逆者追踪）。数字指纹技术多用于网络服务中的版权保护，它主要为那些需要向多个用户提供数字产品，同时希望确保该产品不会被不诚实用户非法再分发的发行商所采用。如果将来发现被非法再分发的数字产品时，发行商可以通过检测其中的“指纹”来追踪该数字产品的原始购买者，正如案件侦破人员可以通过案件发生现场的指纹来追查犯罪嫌疑人。

另外，它们的研究重点不同。数字水印技术的研究重点主要在于设计鲁棒数字水印算法，而数字指纹技术的研究重点主要有三个方面：抗共谋数字指纹编码、指纹复制的有效分发以及非对称数字指纹。

数字水印技术与数字指纹技术之间的关系也很紧密，从大的方面来看，它们都是信息隐藏技术；在实际应用中，数字指纹技术以数字水印技术为基础，指纹的嵌入与提取都需要凭借数字水印算法。

与数字水印技术一样，数字指纹技术也是一种目标明确的被攻击对象，需要有一定的鲁棒性。对于多媒体数据来说，由于数字指纹的嵌入采取了传统的数字水印嵌入方法，所以，也要求数字指纹在单个用户的预谋攻击下能保持较好的鲁棒性。然而，由于通信及网络的高度发达，多个用户很容易联合他们的复制，通过逐个位置比较各自复制，定出部分标识位置进行攻击，这就是所谓的共谋攻击。如果数字指纹系统的嵌入和识别方案设计不合适的话，那么这群共谋者将能成功地制造出一个新的复制，这个复制中已删除了关于他们身份的所有踪迹。因此，共谋攻击对于数字指纹系统来说是一个最大的挑战。设计出能抵抗共谋攻击，并能鉴别共谋攻击者的数字指纹系统显得尤为重要。

4.2.2 数字指纹技术的特性要求

一套完善的数字指纹方案应满足以下几项基本要求。

1. 鲁棒性

数字产品在传输过程中通常会受到各种各样的篡改或信号处理操作，即所谓的**单一复制攻击**（Single Copy Attack）。为了使得数字复制在经过某些不破坏性的修改后仍能够有效地提取到其中的指纹信息，所使用的指纹嵌入算法应具有较高的鲁棒性，要能够抵抗常见的几何失真操作及信号处理操作，使得提取出的信息足以追踪出叛逆者。鲁棒性要求的理想目标是使攻击者无法在不破坏原复制的情况下伪造出一个新的可用复制。

2. 不可感知性（保真性、透明性）

不可感知性是数字指纹技术的一个关键要求。数字作品是供消费者实际使用的，因此它必须具有很好的感知特性。任何视听效果不好的数字产品对于用户而言都是毫无作用的。因此，要求数字指纹在嵌入到数字产品中以后不会引起数字作品视听感知效果的严重失真，必须能够满足用户的正常需求。也就是说与原始复制相比，嵌入指纹后的

数字复制的质量不能有太大的变化。实际上，这是信息隐藏方案的基本要求。

3. 抗共谋攻击

抗共谋攻击是设计数字指纹方案必须要考虑的问题。为了逃避内容提供商的追踪，多个消费者之间可能采取某种线性或者非线性的攻击方式生成盗版的数字复制。在经由共谋攻击所生成的盗版复制中，指纹信息被破坏或者信号强度被削弱，给叛逆者追踪带来严峻的考验。数字指纹方案要求在多用户共谋的情况下，仍然能以较大的概率追踪到至少一名共谋者。通常从以下两个方面考虑共谋容忍性：① 在一定的共谋人数下，发行商能够确定出至少一个叛逆者，该人数称为共谋安全尺寸；② 无论共谋人数的多少（即使超过了上述尺寸），无辜购买者也不能受到指控。

4. 嵌入容量

因为嵌入的内容要实现用户攻击后能留下足够的信息使发行商进行追踪并保证用户 ID 的唯一性，因此要求有足够的嵌入量。

5. 唯一性

不同时间分发给不同用户的不同作品的指纹信息应当是唯一的。用户根据该 ID 可判定该作品版权拥有者的详细信息。

6. 准确性

当发现未授权的数字复制时，不但要能够以较大的概率追踪到至少一个非法的消费者，而且要把无辜消费者受到错误指控的概率控制在很小的范围内。只有这样才能最大限度地保护合法消费者与打击盗版者。

7. 安全性

即使整个嵌入算法公开，由密钥产生的指纹嵌入与提取位置对攻击者仍是未知的，从而增强指纹的抗攻击能力。

8. 效率

要求含指纹拷贝的生成算法和追踪算法的实现具有很好的效率。

9. 其他要求

以上是对数字指纹体制的若干基本要求。此外还有其他一些要求，如实现用户的不可否认性和用户的匿名性等。针对不同需求环境，对指纹体制的各项要求侧重点也会有所不同。在数字指纹体制中，具有较强鲁棒性的指纹嵌入算法，具有抗共谋攻击能力的编码和追踪方案，及有效、快速的协议实现是决定指纹方案的安全性和效率的关键环节。数字指纹技术是一项新的技术，到目前只有 20 年左右的历史。期间，研究人员提出了许多关于数字指纹技术的新观点和新方法。这些研究主要分为以下两个方向：① 研究主要集中于设计抗共谋攻击的数字指纹编码；② 研究检验在不同攻击下的不同数字指纹方案的抗攻击性能。目前，国内外对数字指纹技术的研究主要集中在提高共谋容忍性的指纹编码方案研究、非对称及匿名指纹协议研究、基于多方计算的指纹协议实施效率的提高上。

需要指出，数字指纹技术就其本身而言，仅提供非法使用的检测，并不能避免非法使用，而检测非法使用的能力可以在一定程度上阻止盗版者从事这些非法活动。上面的这些特性要求中的前三条是最关键的，也就是三个容忍分别为，① 共谋容忍：即使攻击者获得了一定数量的拷贝，通过比较这些拷贝，也不应该能找到、生成或者删除该产品的指纹。特别地，根据标识假设，指纹必须有一个共同的交集；② 产品质量容忍：加入

标记不允许明显地降低产品的用途和质量；③ 产品操作容忍：如果攻击者篡改产品，除非因篡改产生太多的噪声以至产品不可用，否则，指纹仍应能存在于产品中。特别地，指纹应能容忍有损数据压缩。第一个“容忍”对于数字指纹系统最为重要，是衡量一个数字指纹系统的核心关键指标之一。Blakley 等人^[49]首先提出了共谋问题，而 Boneh 和 Shaw^[50]首先提出了**嵌入假设**（Marking Assumption），并提出了一种对付多用户共谋攻击的解决方案。这些方案为后来的数字指纹研究工作奠定了坚实的基础。

10. 设计目标

无论共谋者采取什么样的攻击手段，数字版权所有者总的目的就是捕获共谋者，停止非法拷贝。但是在不同的情况下侧重点将有所不同，通常数字指纹设计的目标如下。

（1）捕获一个共谋者

这个设计需要指纹系统在错误指控一个无辜用户的可能性最小化的基础上，使捕获一个共谋者的概率最大。基于此设计目的的性能评价标准失败的可能性有两种：一种是系统没有识别出任何一个共谋者，即**漏检**（False Negative）；第二种是系统将一个无辜用户指证为共谋者，即**虚警**（False Positive）；

（2）捕获部分共谋者

这个设计是捕获尽可能多的共谋者，尽管这样可能指控更多的无辜用户。此设计目的的性能评价标准是成功捕获的共谋者的期望值和错误指证无辜用户的期望值；

（3）捕获全部共谋者

此设计使捕获全部共谋者的概率尽可能大。当然，也要求指控无辜用户的可能性应该保持在一个合理的范围内。

4.2.3 数字指纹技术的分类

1. 按照协议类型分类

根据数字指纹协议，数字指纹技术属于密码学范畴，在应用中属于电子交易、电子商务和公平交易问题，包括分发、追踪、认证和仲裁等方面；从理论上讲它属于两方/多方安全计算的密码学问题。1996 年德国学者 Pfitzmann 和 Schunter 首先将密码理论引入到数字指纹之中，将数字指纹分为**对称指纹**（Symmetric Fingerprinting）^[51]和**非对称指纹**（Asymmetric Fingerprinting）^[52]两类，继而在非对称指纹中又提出**匿名指纹**（Anonymous Fingerprinting）^[53]。

（1）对称指纹

由销售商生成指纹及标记信息，并嵌入于载体作品中，当销售商发现非法拷贝时，依靠拷贝中的信息能够找到叛逆者（盗版者）。销售商与购买者都具有含指纹拷贝，但只有销售商知道指纹信息。

（2）非对称指纹

只有购买者拥有含指纹拷贝，销售商不知道指纹的全部信息，避免销售商进行欺骗，当销售商发现了非法拷贝时，能通过提取指纹并核对销售记录找出盗版作品的购买者。

（3）匿名指纹

在非对称指纹的基础上，购买者在不泄漏身份信息情况下实现数字产品购买，销售商得到盗版拷贝后，通过权威机构或者登记中心，在不需要了解购买者的隐私信息基础上识别出叛逆者。

对称指纹通常不包含指纹协议，因此早期的指纹编码也被称作对称指纹。因为销售

商完全掌握指纹及其拷贝, 存在诬陷无辜购买者的可能, 所以它不能作为法庭上的证据对盗版用户进行控告。非对称指纹协议与匿名指纹协议就是针对这一问题提出的, 并与互联网环境下的电子商务和电子货币紧密相关, 以实现指纹生成与追踪的自动化、法律化, 逐渐形成数字版权销售、认证、监视、追踪及仲裁在内的完善的版权管理系统。

2. Wagner 给出的分类

数字指纹技术可以按照加入指纹的产品、检测灵敏度、嵌入指纹的方法和生成的指纹等特征进行分类。Wagner 给出了如下的分类方法^[48]:

(1) 基于加入指纹的产品进行分类

基于加入指纹的产品进行分类, 指纹可分为数字指纹和物理指纹。如果加入指纹的产品是数字格式, 我们称之为数字指纹。如果一个产品能用自己的物理特性与其他产品区分开来, 我们称之为物理指纹。

(2) 基于对侵害的检测灵敏度进行分类

基于对侵害的检测灵敏度进行分类, 可分为完美指纹、统计指纹和门限指纹。如果对产品的任何修改使得指纹不可识别的同时, 也导致了产品不可用, 我们称此指纹为完美指纹。统计指纹指假定有足够多的误用产品可供检测, 指纹识别器能以任意期望的可信度来确认非法用户, 当然这种识别器不是绝对可靠的。门限指纹是上述两种指纹类型的混合, 它允许一定程度的非法使用, 只有达到门限值时才启动非法拷贝识别, 追踪拷贝者。

(3) 基于嵌入指纹的方法进行分类

基于嵌入指纹的方法进行分类, 可分为识别指纹、删除指纹、添加指纹、修改指纹等。

(4) 基于生成的指纹进行分类

基于生成的指纹可分为离散指纹和连续指纹。如果生成的指纹是有限的离散取值, 那么就称该指纹是离散的, 例如数字文件的哈希数列。如果生成的指纹是有限的连续取值, 那么就称该指纹是连续的, 大部分物理指纹都属于这种类型。

4.3 数字指纹系统模型和性能评价

4.3.1 数字指纹系统模型

数字指纹体制主要由两部分构成, 一是用于向拷贝中嵌入指纹并对带指纹拷贝进行分发的指纹分发体制; 另一部分是实现对叛逆者进行追踪并仲裁的追踪体制。上述两种体制通过发行商、用户(还可能有**登记中心**、**仲裁者**等实体)之间的一系列协议实现, 因此数字指纹体制也可以分为**算法**和**协议**两部分。其中, **算法**包括指纹的编码和解码、指纹的嵌入和提取以及拷贝的分发策略等内容, 而**协议**部分则规定了各实体之间如何进行交互以实现具有各种特点的拷贝分发和追踪体制(如实现用户的匿名性等)。数字指纹体制的简单模型如图 4.2 所示。其中, 用户 j 的信息 m_j 由用户提供或由其与发行商(或登记中心等实体)通过一系列交互后生成。它通常包括用户的身份信息及该次购买过程的描述信息。有关用户 j 的信息将被按照一定规则进行编码并嵌入到发行商要出售的原拷贝 c 中。用户直接得到含指纹拷贝 s_j 或由发行商将含指纹拷贝 s_j 发放给用户, 同时发行商和用户得到有关交易记录。不诚实的用户可能会直接分发他所得到的拷贝, 也可能与其他用户联合获得新的拷贝后分发。无论是哪种情况, 非法分发的拷贝 s' 中都会留下参与非法活动用户的指纹信息。一旦发行商发现了非法拷贝, 他将运用相应的指纹提取及指纹解码技术, 并运用追踪算法追踪叛逆者。一般来讲, 只要发行商能够成功地追踪

出一个叛逆者，就认为该追踪算法是成功的；如果该追踪过程不能给出一个共谋者或者将一个无辜用户认为是叛逆者，则认为该追踪算法是失败的。指纹的编码方法以及共谋人数大小是影响追踪成败的关键因素。

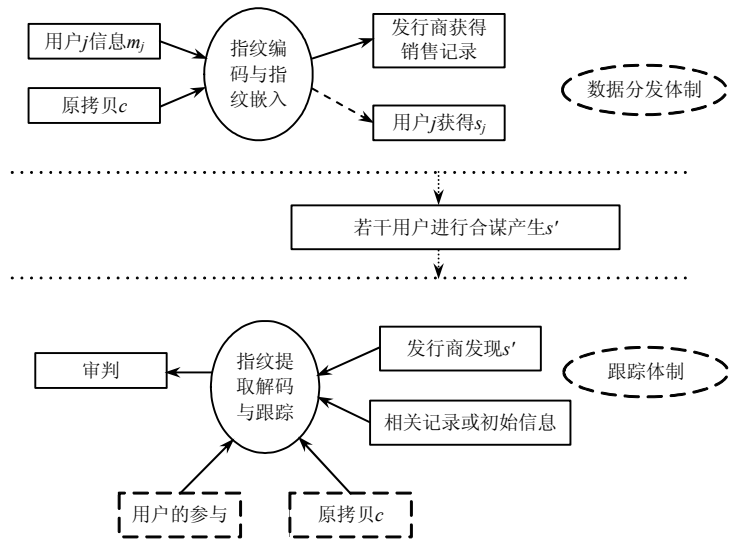


图 4.2 数字指纹系统模型

4.3.2 数字指纹系统的攻击手段

数字指纹攻击手段和数字指纹系统抗攻击方法是矛盾的对立统一的两个方面。不断地分析和研究指纹攻击手段，有助于我们更加深入地发现并认识指纹系统的缺点和不足之处，从而改进算法、协议及系统设计方法，进一步完善和提高系统的性能。

数字指纹系统可能受到的攻击主要分为两类：一类是单个用户对其含指纹拷贝进行信号处理操作或恶意攻击，以删除指纹或将指纹变成其他用户的指纹，这样的攻击称为单用户攻击，这类攻击与数字水印系统遭受的攻击相同；另一类是多个用户的共谋攻击手段，由于数字指纹系统为每个用户分配一个唯一的指纹，从而每个用户获得的含指纹作品存在细微的差别，几个用户联合起来可产生一个新的含密作品，这里的秘密信息可能是一个没有参与共谋者的指纹，也可能是一个乱码。下面分别介绍这两类攻击。

1. 单用户攻击

单用户攻击是指单个用户通过对其拥有的含指纹产品进行各种操作，以期去掉数字产品中的指纹或者将指纹转换成其他用户的指纹，这时数字指纹系统和数字水印系统所遭遇的攻击基本相同，一般来说攻击方法可以分为三类：**去除攻击**（Removal Attack）、**同步攻击**（Synchronization Attack）、**协议攻击**（Protocol Attack）。

（1）去除攻击

去除攻击也称为稳健性攻击，其目的是从数字产品中去除掉指纹，并且不影响数字产品的正常使用。这类攻击大体可分为两类：信号处理攻击和分析攻击。

信号处理攻击包括一些基本的信号处理操作，这主要是针对整个数据的处理。针对于不同的数据，有不同的稳健性攻击方式，例如针对视频文件和图像，常采用一些图像处理的方法删除指纹，如 JPEG 压缩、低通滤波、加噪、旋转、剪切、尺度变换等；针

对音频文件采用剪切、滤波、MP3 压缩、重采样、重量化等。对于这类攻击，数字指纹技术主要通过数字指纹嵌入过程中提高嵌入指纹的稳健性来对指纹加以保护。

分析攻击也称直接攻击，包括在指纹的嵌入和检测提取阶段采用特殊方法来擦除或减弱产品中的指纹。

(2) 同步攻击

同步攻击也称为表达攻击，它不需要利用算法来去除数字产品中的指纹，而是试图通过对含指纹载体做各种修改以使得检测器失效，破坏了载体与指纹的同步性。这类攻击主要包括几何失真攻击 (Geometrical Distortion Attack)、马赛克攻击 (Mosaic Attack)、抖动攻击 (Jitter Attack)、Oracle 攻击。其中，几何失真攻击由于简单易行，是较为常见的一类攻击，主要包括旋转 (Rotation)、缩放 (Scaling)、平移 (Translation)、剪切 (Cropping) 和图像反转等。其中，旋转、缩放、平移简称为 RST。

(3) 协议攻击

协议攻击使检测器的结果错误或不明确，从而不能唯一地确定版权的所有，引起版权纠纷。协议攻击主要包括**解释攻击** (Interpretation Attack) 和**拷贝攻击** (Copy Attack)。解释攻击又称为死锁攻击，它通过伪造假原始数字产品或含指纹产品来制造混乱，从而任何人可以声明其对产品的所有权，导致版权争议。

拷贝攻击是从含指纹作品中估计出指纹，并将估计的指纹嵌入到其他数字产品，破坏版权产品的合法性。拷贝攻击的目的不是破坏水印，而是像重调制攻击一样从含指纹媒体中估计指纹，然后把它复制到目标数据中生成伪装媒体，复制指纹的过程既不需要知道指纹算法，也不需要密钥。

2. 共谋攻击

由于每个授权用户的含指纹产品都有微小的差异，部分用户可以联合起来篡改数据以达到去除指纹或陷害无辜用户的目的，我们称之为共谋攻击。共谋攻击的成功与否，不仅在于攻击后产生的新数字产品的变化程度，还在于数字指纹系统的追踪能力。

研究数字指纹系统遭受共谋攻击时，一般以 Boneh 和 Shaw 提出的嵌入假设^[50]为前提，即：假定各个指纹拷贝的嵌入位置相同。假设数字指纹系统的用户数为 U ，载体数据中供指纹嵌入的 N 个位置上的数据值构成向量 $c=(c(1), c(2), \dots, c(N))$ ，发行商为每个用户 u_i 产生一个唯一的长度为 N 的指纹 $m_i=(m_i(1), m_i(2), \dots, m_i(N))$ ， $i=1, 2, \dots, U$ 。则嵌入指纹 m_i 的含指纹作品中被修改的 N 个位置上的数据 $s_i=(s_i(1), s_i(2), \dots, s_i(N))$ 可按下式计算

$$s_i(j)=c(j)+\alpha \cdot m_i(j) \quad (4.1)$$

其中， α 表示嵌入强度。假设 U 个用户中有 L 个用户参与共谋，共谋者的索引集合为 $\Gamma=\{i_1, i_2, \dots, i_L\}$ ，共谋者的含指纹作品数据向量集合为 $\{s_l, l \in \Gamma\}$ ，共谋产生的新拷贝记为 s' 。

根据这些设定条件，可以定义如下八种具体的共谋攻击方式。

(1) 均值攻击

新拷贝 s' 的每个嵌指纹位置的数据是共谋者该嵌指纹位置对应数据的平均，即

$$s'(j)=\sum_{i \in \Gamma} \frac{s_i(j)}{L}, \quad j=1, 2, \dots, N \quad (4.2)$$

(2) 最小值攻击

新拷贝 s' 的每个嵌指纹位置的数据取共谋者该嵌指纹位置对应数据的最小值，即

$$s'(j)=\min_{i \in \Gamma} \{s_i(j)\}, \quad j=1, 2, \dots, N \quad (4.3)$$

(3) 最大值攻击

新拷贝 s' 的每个嵌指纹位置的数据取共谋者该嵌指纹位置对应数据的最大值, 即

$$s'(j) = \max_{i \in \Gamma} \{s_i(j)\}, \quad j = 1, 2, \dots, N \quad (4.4)$$

(4) 中值攻击

新拷贝 s' 的每个嵌指纹位置的数据取共谋者该嵌指纹位置对应数据的中值, 即

$$s'(j) = \text{median}_{i \in \Gamma} \{s_i(j)\}, \quad j = 1, 2, \dots, N \quad (4.5)$$

(5) 最小最大值攻击

新拷贝 s' 的每个嵌指纹位置的数据取共谋者该嵌指纹位置对应数据的最小值和最大值的平均, 即

$$s'(j) = \frac{1}{2} (\min_{i \in \Gamma} \{s_i(j)\} + \max_{i \in \Gamma} \{s_i(j)\}), \quad j = 1, 2, \dots, N \quad (4.6)$$

(6) 改良负攻击

新拷贝 s' 的每个嵌指纹位置的数据取共谋者该嵌指纹位置对应数据的最小值+最大值-中值, 即

$$s'(j) = \min_{i \in \Gamma} \{s_i(j)\} + \max_{i \in \Gamma} \{s_i(j)\} - \text{median}_{i \in \Gamma} \{s_i(j)\}, \quad j = 1, 2, \dots, N \quad (4.7)$$

(7) 随机负攻击

新拷贝 s' 的每个嵌指纹位置的数据分别以概率 p 和 $1-p$ 取共谋者该嵌指纹位置对应数据的最小值和最大值, 即

$$s'(j) = \begin{cases} \min_{i \in \Gamma} \{s_i(j)\} & \text{以概率 } p \\ \max_{i \in \Gamma} \{s_i(j)\} & \text{以概率 } 1-p \end{cases}, \quad j = 1, 2, \dots, N \quad (4.8)$$

(8) 拷贝粘贴攻击

它通过剪切每个共谋者的数字产品拷贝 s_i 的不同部分, 并将它们粘贴在一起, 从而构成新的拷贝 s' 。

4.3.3 数字指纹技术的性能评价

目前信息隐藏的研究多集中于设计满足不可感知性的信息隐藏算法, 而对信息隐藏的本质缺少系统性的研究, 造成信息隐藏技术的评价缺乏全面、客观和统一的标准。只有从研究信息隐藏的本质性能出发, 才能真正适应信息隐藏实际应用和技术发展的要求, 发现信息隐藏算法的脆弱性和局限性, 改进与完善信息隐藏算法。因此, 探索更加全面、客观的信息隐藏性能评价标准具有十分重要的意义。下面从六个方面介绍数字指纹系统的性能评价问题, 其中鲁棒性、透明性、容量、防利用性能和可靠性这五个方面适用于数字水印技术和数字指纹技术, 而抗共谋攻击能力专门针对数字指纹技术。

1. 数字指纹技术的鲁棒性评价

鲁棒性是数字水印系统和数字指纹系统都应该满足的重要特性之一。目前, 没有任何一种方法来对某个水印嵌入系统的鲁棒性进行数学证明。最常用的评价做法是: 能够经受住现有鲁棒性攻击的水印算法或指纹算法就是鲁棒性好的算法。因此, 数字水印技术和数字指纹技术的鲁棒性评价是建立在相应的攻击基础之上的。关于鲁棒性我们在第3章已经介绍过, 这里只是给出两个常用指标: 相关性和比特错误率。

(1) 相关性度量

在数字水印系统和数字指纹技术中通常使用 NC (Normalized Correlation) 系数作为提取出来的水印 (指纹) w' 与原水印 (指纹) w 的相似程度的客观衡量, 计算公式如下

$$NC = \frac{\sum_{i,j} w(i,j) \cdot w'(i,j)}{\sqrt{\sum_{i,j} w(i,j)^2} \cdot \sqrt{\sum_{i,j} w'(i,j)^2}} \quad (4.9)$$

(2) 比特错误率

BER (Bit Error Rate) 表示提取出的错误比特数占全部嵌入比特数的百分比, 计算公式如下。

$$BER = \frac{100}{N} \sum_{i=1}^N b(i), \quad b(i) = \begin{cases} 1 & w'(i) \neq w(i) \\ 0 & w'(i) = w(i) \end{cases} \quad (4.10)$$

其中, N 为提取的水印序列长度。

通常在只需判断水印有无或只能通过相关性计算来检测水印信息的场合下, 使用相关性度量; 而在可以提取出水印信息的场合下, 使用 BER 度量。

2. 数字指纹技术的保真度 (不可感知性) 评价

不可感知性是信息隐藏的首要条件之一。目前对不可感知性的评价, 主要提出了两种评价方法: 主观评价和客观评价。以图像为例, 主观评价是依据人类视觉对失真和畸变进行主观评价, 如对伪轮廓、方块效应、振铃效应、噪声颗粒和几何失真等分辨图像的失真和畸变程度。然而, 经验不同的人员对图像的主观评价结果相差很大。客观评价不依赖于主观感觉, 可以使基于不同原理信息隐藏方法按照同一个固定的评价标准进行计算, 互相间比较的结果也更趋于合理。常用的基于像素的图像视觉失真度量标准有很多, 如信噪比、峰值信噪比、拉普拉斯均方误差、归一化互相关、全局西格马信噪比、直方图相似性等。上述这些内容在第 2 章和第 3 章中已经详细讨论, 在此不再赘述。

3. 数字指纹技术的容量性能评价

衡量通信系统性能的一个关键指标为信道容量, 对于数字水印系统, 信道容量是指当存在攻击时, 一幅数字作品所能加载的最大信息量。为保证通信的价值, 在不可感知性和载体一定的前提下, 应尽量在载体中嵌入更多的信息。隐写理论起源于 Simmons 的认证体系和 Shannon 的通信保密系统, 其基本理论是熵理论。假设 c 是载体对象, m 是秘密信息, s 是隐写对象, 由于 $H(m) \leq H(c_p | s)$, 因此, 即使隐写分析者知道了 s 和 c_p (一部分 c) 的统计特性, 也无法知道秘密信息 m , 这就给出了隐藏容量的上界。评价水印嵌入系统时, 首先根据其算法建立容量分析模型, 再在相应的模型下, 不断增加容量直到达到理论上的最大安全容量估计值。如果此时水印嵌入软件的最大容量等于或小于该模型的最大安全容量, 那么可以认定该算法在防容量估计方面是安全的。

4. 数字指纹技术的防利用性能评价

当前隐写嵌入系统通常分为三类: 纯隐写嵌入、对称密钥隐写嵌入和非对称密钥隐写嵌入。隐写嵌入软件防止中间人攻击的关键环节在于密钥的分发机制, 纯隐写嵌入在隐秘通讯之前不需要进行秘密信息交换, 整个系统的安全性完全依赖于系统本身; 对称密钥隐写嵌入类似于对称密钥密码算法; 非对称密钥隐写嵌入就象公钥密码学一样, 不依赖于秘密密钥的交换。在防利用方面安全等级最高的是非对称密钥隐写嵌入技术。

5. 数字指纹技术的可靠性评价

可靠性表征数字水印/指纹算法的可靠程度。可用鲁棒性—攻击强度曲线，鲁棒性—保真度曲线或攻击强度—保真度曲线等表示。由于算法鲁棒性与保真度、容量等因素有关，为了能够进行合理的性能评估，通常固定某些因素。例如鲁棒性—攻击强度曲线反映了在固定保真度和容量的前提下，水印/指纹算法鲁棒性与攻击强度之间的函数关系。为避免采用一个固定阈值比较不同的水印/指纹算法可能导致的错误评价结果，可引入接收者操作特征曲线（ROC, Receiver Operating Characteristic Curve）采用不同的判定阈值来进行比较。ROC 曲线对整体水印/指纹方案的性能与可靠性评估有十分重要的作用。有关 ROC 的介绍可参见第 7 章 7.2.2 节。

6. 数字指纹技术的抗共谋攻击能力评价

数字指纹系统除了需具备水印系统的一些特性之外，其最显著的特征是其抗共谋攻击的能力。因此，抗共谋攻击能力是衡量数字指纹技术性能的特殊重要指标。在分析设计抗共谋系统性能方面，一部分人把着眼点放在研究某些数字指纹系统或方案对于不同种类的共谋攻击的抵抗性能上，主要通过考查几个重要的量（需要嵌入的信息长度 N ，指纹系统中总的用户数 U 和参加共谋的敌手数量 L ），找出它们的联系或者在某些给定条件下的边界情况等。另一部分人从载体的角度出发，利用相关性、可疑载体特性和载体标准特性来研究系统抵抗共谋攻击的能力。

假设指纹数据库由 U 个相互独立的指纹向量组成 $\{c_i\}$, $i \in [1, U]$ ，共有 L 个用户参与了共谋，共谋者索引集合为 $\Gamma = \{i_1, i_2, \dots, i_L\}$ 。 T_i 为第 i 个指纹与提取指纹的相似度， h 为阈值，则检测概率、漏警概率、虚警概率、部分用户检测概率与部分用户虚警概率分别用概率模型表示如下^[54]。

① 检测概率 P_d : 识别至少一个共谋者的概率。

$$P_d = P\left[\max_{i \in \Gamma} T_i > h\right] \quad (4.11)$$

② 漏警概率 P_{fn} : 识别不到任何一个共谋者的概率。

$$P_{fn} = P\left[\min_{i \in \Gamma} T_i \leq h\right] \quad (4.12)$$

③ 虚警概率 P_{fp} : 把一个无辜者误认为共谋者的概率。

$$P_{fp} = P\left[\max_{i \notin \Gamma} T_i > h\right] \quad (4.13)$$

④ 部分用户检测概率 F_d : 识别一部分用户的概率。

$$F_d = E[P[T_{i \in \Gamma} > h]] \quad (4.14)$$

⑤ 部分用户虚警概率 F_{fp} : 把一部分无辜者误认为共谋者的概率。

$$F_{fp} = E[P[T_{i \notin \Gamma} > h]] \quad (4.15)$$

尽管数字指纹系统总的目标是追踪共谋者，但是根据不同的情况及版权者要达到的目的，追踪目标也有所区别。根据追踪共谋者数量不同大致可分为三种：识别一个（Catch One）、识别多个（Catch Many）、识别所有（Catch All）^[55]。

（1）识别一个

在这种情形中，指纹系统的目标是以最大的概率追踪到至少一个共谋者，同时使得无辜用户误判为共谋者的错误概率最小。如果指纹系统识别不到共谋者或者把一个无辜者误认为共谋者表示则检测失败，因此指纹系统要求

$$P_d \geq \lambda_d \quad \text{且} \quad P_{fp} \leq \lambda_{fp} \quad (4.16)$$

其中 λ_d 、 λ_{fp} 是根据系统需求预先设定的参数。识别一个共谋者的检测标准适用于向法庭提供证据的情形。

(2) 识别多个

在这一应用下，系统的目标是识别尽可能多的共谋者，同时应使把无辜用户误判为共谋者的概率尽可能小。系统同时要满足

$$F_d \geq \lambda'_d \quad \text{且} \quad F_{fp} \leq \lambda'_{fp} \quad (4.17)$$

其中 λ'_d 、 λ'_{fp} 是根据实际需求预先设定好的参数。识别多个共谋者的检测标准用于先确定可疑用户，然后再寻求其余证据。

(3) 识别所有

在这一应用下，系统的目标是识别所有共谋者，同时应使把无辜用户误判为共谋者的概率控制在可以容忍的范围内。性能标准由检测效率 R 和识别所有共谋者的概率 $P_{d,all}$ 来衡量，分别定义如下：

$$R = \frac{(U - L) \times F_{fp}}{L \times F_d} \quad (4.18)$$

$$P_{d,all} = P[\min_{i \in I} T_i > h] \quad (4.19)$$

其中， U 表示用户总数， L 表示共谋者总数。系统要求

$$R \leq \theta_r \quad \text{且} \quad P_{d,all} \geq \theta_d \quad (4.20)$$

θ_r 、 θ_d 是根据需要预先设定好的参数。这一检测目标主要用于重要数据安全保密方面，稍有疏漏就会引起重大损失的场合，例如军事机密方面的保护。

4.4 指纹编码和指纹协议概述

由于数字指纹系统里采用的嵌入方法可以借鉴数字水印中的嵌入技术，而且近年来已出现了各种各样的具有一定鲁棒性的嵌入算法，故本章对数字指纹嵌入方法不作为讨论重点。本章主要关注指纹编码（以及相应的追踪算法）和指纹协议，因为它们是近年来国际上对数字指纹的研究热点。本节首先对这两个方面进行概述，从下一节开始将以常见的指纹协议为线索对数字指纹技术加以逐步介绍。

4.4.1 指纹编码概述

指纹编码是指在一定的假设下，将获得的与用户有关的信息按照一定的规则进行编码，生成具有一定抗攻击能力的码字的过程。指纹编码要不仅能够成功的对单个盗版行为进行追踪，还要考虑多个用户进行共谋的情况，对多个参与盗版的用户行为进行追踪。

由于数字指纹方案要对抗用户的共谋攻击，通常发行商会对用户的指纹进行编码，以增加该指纹方案的共谋容忍能力，这种编码称为共谋容忍编码。若一个数字指纹体制能够抵抗共谋攻击，则称该指纹编码方案是**共谋安全的**（Collusion-secure）。本小节着重讨论指纹的共谋容忍编码问题。指纹的共谋容忍编码通常包括两个部分：指纹的编码算法和追踪算法。指纹编码方案是指：在一定假设下，将获得的与用户有关的信息按照一定的规则进行编码，生成具有一定抗攻击能力的码字的过程；追踪方案则是指：当发行商获得盗版拷贝时，运用一定的解码规则判断出叛逆者的过程。好的指纹编码和追踪算法是发行商能正确追踪到叛逆者的关键因素，每一个指纹编码方案都有相应的追踪体

制。从追踪成功的概率来讲, 指纹编码方案可以分为确定性追踪和概率性追踪方案; 从码字的分布而言, 可以分为连续指纹方案和离散指纹方案; 此外, 从码字是否随机来讲, 还可以分为随机指纹方案和利用某些特殊的组合结构构造的指纹编码方案。现有的指纹编码方案主要是概率性追踪方案, 通过使用连续的或离散的随机信号实现指纹的编码, 下面对其进行简要介绍。

1. 连续指纹方案

在此类编码方案中, 用户码字中的每一个码元取自一个连续的集合, 如一个实数区间, Cox 等于 1997 年在文献[56]中提出的方案是此类编码方案的典型代表。在该文中, Cox 等用独立随机的正态采样序列作为要嵌入的水印信息, 当用作指纹时, 为每个用户选取不同的采样序列, 序列间是独立的。这里指纹的取值不限于离散的整数值, 而是服从正态分布 $N(0,1)$ 的随机实数序列 m 。在追踪时, 发行商从非法拷贝中提取出嵌入信息 m' , 将其与 m 做相关检测, 如果相关值大于某一个门限值, 则认为非法拷贝中含有该指纹 m 。我们称这种体制是 CKLS 体制 (取自四位作者的姓氏的第一个字母)。该文中还给出具体的指纹嵌入方法, 它是一种基于变换域的嵌入方法: 对整个图像进行离散余弦变换 DCT, 然后将嵌入内容叠加在 DCT 域中幅值最大的前 N 个系数上 (不包括直流分量), 通常为图像的低频分量。数字指纹序列记为 $m=\{m_i\}$, $i=1, 2, \dots, N$, 记所选择的 N 个系数为 $C=\{C_i\}$, $i=1, 2, \dots, N$, 则嵌入算法为 $C'_i=C_i(1+\alpha m_i)$, 其中常数 α 为尺度因子, 控制信息嵌入的强度, 其大小正比于相应频率分量的信号强度 (为简单起见, 对所有系数用同一嵌入强度)。然后用新的系数做反变换得到含水印图像。在提取时, 分别计算原始图像和含水印图像的 DCT 变换系数便可得到嵌入的水印。该算法不仅具有较好的保真性, 而且有较强的鲁棒性, 可抵抗有损的 JPEG 压缩、滤波、D/A 和 A/D 转换及重新量化等信号处理, 也可经受一般的几何变换如剪切、缩放、平移及旋转等操作, 对复印、扫描等处理也具有较强的鲁棒性。作者对于共谋者的平均攻击也进行了讨论, 实验显示: 当共谋者采用平均攻击生成盗版拷贝时, 从盗版拷贝中提取出的信息与共谋者指纹的相关性明显高于与无关指纹的相关性, 即 CKLS 方案具有较好的共谋容忍性。

到了 1998 年, Kilian 等^[57]为 CKLS 体制构建了明确的数学模型: 设指纹信号是 N 维向量 m (其分量独立取自于正态分布 $N(0, \sigma^2)$), 设原拷贝是 N 维向量 c , 指纹嵌入过程是 $c'=c+m$, 水印检测方法就是 CKLS 体制中的相关检测法。该模型主要基于以下 3 个假设。① 代表原拷贝的向量 c , 其分量独立取自于正态分布 (不失一般性, 假设为 $N(0, 1)$); ② 假设攻击者不知道水印的检测器, 即他不能用水印检测器检验自己的攻击效果; (3) 原拷贝 c 与嵌入信息后的拷贝 s 的相似性准则为 $\|c-s\|_2 \leq \delta \sqrt{N}$ (其中, N 为序列长度, $\|\cdot\|_2$ 表示欧几里德范数)。发行商和攻击者产生的可用拷贝必须满足相似性准则。Kilian 等指出上述方案是 \mathcal{C} -安全的, 其中 $\mathcal{C} = O(\sqrt{N/\ln U})$, U 是不同指纹的个数, 即 \mathcal{C} 个共谋者生成的拷贝 c^* , 要么与 c 不满足相似性准则, 要么发行商可以从共谋拷贝中检测出至少一个共谋者, 且无辜用户以较高的概率不受共谋者的陷害。实际上, c 的各个分量的独立同分布特性不一定能够得到保证, 因此文献[58]对假设①进行了放宽, 并得出 $O(\sqrt{N/\ln N})$ 个共谋用户可以使任何 CKLS 型的方案不再是安全的, 其中 N 是拷贝的向量维数。值得注意的是, 上述方案在对用户进行追踪时的计算复杂度是 $O(U)$ (其中 U 是用户的总数目)。当用户数目较大时, 追踪效率较低。因此, 文献[59]对 CKLS 型指纹编码方案的追踪算法效率进行了改进。作者利用 Reed-Solomon 纠错编码, 并采用级联码的思想, 以 CKLS 编码作为内码, 构造了改进的 CKLS 方案。该方案以适当减小共谋

尺寸为代价, 使指纹追踪算法的效率从与用户数目 U 成线性关系提高到与 $\log U$ 成多项式关系。以上利用服从正态分布的随机序列作为指纹的方法具有较好的共谋容忍性, 不过还可以将 CKLS 模型中的三个假设进行放宽进行深入讨论。

2. 离散指纹方案

离散指纹方案的经典文献是 Boneh 和 Shaw 所提出的方案^[50], 该文中的编码方法与使用的数据嵌入算法无关, 只要其能够满足“嵌入假设”(Marking Assumption)。基于该假设, 运用级联码的思想并在编码方案中引入随机性, 给出了一种指纹码字长度 N 与用户数目 U 的对数及共谋容忍尺寸 C 的四次方成正比的指纹编码方案 ($N=O(C^4 \log(U/\varepsilon) \log(1/\varepsilon))$), 其中, ε 含义是发行商能以大于 $1-\varepsilon$ 的概率进行追踪)。该编码及追踪思想在后来的多篇关于指纹体制构造的文献中得到应用。因其编码思想较为经典, 对其简述如下。

嵌入假设: 共谋用户通过对比他们的拷贝, 只能在拷贝相异之处发现指纹并进行修改。对于没有发现不同的指纹所在之处, 除非将拷贝变得无用, 他们不能对该处的指纹进行修改。对于通过对比拷贝发现的指纹比特, 假设共谋者只可能将其修改为 $\{0, 1, ?\}$ 这 3 种状态 (?表示无法识别是 0 还是 1)。

Boneh 和 Shaw 首先给出了带有错误概率 ε 的 n -安全的 (l, n) 码。码本记为 $\Gamma_0(n, d)$, 令 e_i 是高度为 n 的一个列, 其中前 i 个元素为 1, 其余为 0。则 $\Gamma_0(n, d)$ 包括所有列 e_1, e_2, \dots, e_{n-1} , 每个重复 d 次。 d 将决定错误概率 ε 。对上述列进行排列后的每一行称为一个码字, 共 n 个码字, 码长为 $l=d(n-1)$ 。 e_i 所占的 d 个比特位记为 B_i 。例如, 针对 4 个用户的码本 $\Gamma_0(4, 3)$ 为:

	$e_1 e_1 e_1$	$e_2 e_2 e_2$	$e_3 e_3 e_3$	
a_1	111	111	111	
a_2	000	111	111	(4.21)
a_3	000	000	111	
a_4	000	000	000	

当发行商分配码字时, 它并不是简单地按顺序将 a_i 分配给第 i 个用户 u_i , 而是随机地选择一个置换 π (π 对用户是保密的), 然后将 $\pi(a_i)$ 作为第 i 个用户的指纹, 称上述编码方案为带有置换 π 的 $\Gamma_0(n, d)$ 编码方案。在文献[50]的定理 12 中作者指出了: 对于 $n \geq 3$, 且 $\varepsilon > 0$, 令 $d=2n^2 \log(2n/\varepsilon)$, 则带有置换 π 的 $\Gamma_0(n, d)$ 是带有错误概率 ε 的 n -安全 (l, n) 码。因为上述的码本 $\Gamma_0(n, d)$ 中码长为 $l=d(n-1)$, 显然当数字产品的发行量较大时, 该方案很不实用。

基于 $\Gamma_0(n, d)$, 设用户数是 U , Boneh 和 Shaw 给出了一种码长正比于 $[\log U]^{O(1)}$ 的 C -安全编码方案, 其中 $C=O(\log U)$ 。其基本思想是用 n -安全 (l, n) 码的码本作为字母表, 基本构造如下: 设 A 是 (L, U) 码, 其码元独立随机地取自尺寸为 n 的码表。将 $\Gamma_0(n, d)$ 作为该码表, 得到的新码本记为 $\Gamma_1(L, U, n, d)$ 。设 $a' \in \Lambda$, $a'=a_1 a_2 \dots a_L$, 若 a_i 取原码表中第 t_i 个码字, 则新码本中相应码字的相应码元取自 $\Gamma_0(n, d)$ 中第 t_i 个码字。注意, 嵌入 L 个 $\Gamma_0(n, d)$ 中的码字时, 需要使用 L 个置换, 这 L 个置换对用户是保密的。此外, 码本 A 对用户也是保密的。文献[50]中的定理 17 指出了: 给定正整数 U , C 和 $\varepsilon > 0$, 令 $n=2C$, $L=2C \log(2U/\varepsilon)$, 且 $d=2n^2 \log(4nL/\varepsilon)$, 则 $\Gamma_1(L, U, n, d)$ 是带有错误概率 ε 的 C -安全码。该码本中有 U 个码字, 每个码字的长度为 $O(Ldn)=O(C^4 \log(U/\varepsilon) \log(1/\varepsilon))$ 。可见, 当 $C=O(\log U)$ 时, 码长为 $[\log U]^{O(1)}$, 这大大增加了该方案的实用性。

尽管文献[50]中的编码方法具有较好的共谋容忍性, 但该文中的追踪算法存在一个

较大的问题,即一个误发的随机错误就可能导致错误判断;若参与共谋的用户所对应的码字标号有较大间隔,则会以较高概率产生误判。为此,后期研究将文献[50]中的假设予以放宽(如允许在不可侦察位以一定概率出现错误码元)或者提出更切实际的嵌入假设。总的来看,后期对这类编码方案的讨论主要集中于在一定的共谋容忍尺寸下,如何降低用户的码长并尽量放宽嵌入假设以降低对嵌入算法强度的要求,进一步改进追踪算法的效率。此外,也有一些学者提出直接利用随机二进制码元作为指纹的若干体制,编码思想较为简单,但码字往往较长。利用某些具有特殊组合性质的二进制(或多进制)码字对指纹编码进行研究,一直是指纹编码研究的热点之一。例如,基于对偶二元汉明码的抗两个人共谋攻击的指纹编码方案、基于有限几何的指纹编码方案、基于 BIBD(平衡非完全区组设计)构造的指纹编码方案。还有些学者对 IPP 码(Identifiable Parent Property,可确认父元码)、FP 码(Frameproof,防陷害码)、SFP 码(Secure FP,安全防陷害码)、TA 码(Tracibility,可追踪码)的组合特性及他们的相互联系进行了研究;同时运用组合论和编码理论中的若干方法,讨论了有关码字结构中参数界的问题,并给出了几种关于 IPP 码,TA 码等的构造方法。总之,由于要对抗共谋攻击,需要对指纹进行一些特殊的编码,同时对要嵌入信息的长度也提出了更高的要求。如何结合嵌入技术的发展状况设计较短的实用的码字,仍将是指纹编码中的核心课题。针对抗共谋攻击的指纹方案的详细讨论见 4.9 节。

4.4.2 指纹协议概述

我们已经知道,数字指纹体制主要由两部分构成。一部分是用于向拷贝中嵌入指纹并对带指纹拷贝进行分发的拷贝分发体制;另一部分是实现对叛逆者进行追踪并仲裁的追踪体制。往往上述两部分通过发行商、用户(还可能有登记中心、仲裁者等实体)之间的一系列协议实现。协议部分则规定了各实体之间如何进行交互以实现具有各种特点的拷贝分发和追踪体制。数字指纹协议经历了对称、非对称、匿名三个发展阶段,其中对称指纹方案不能提供不可否认性;非对称方案虽能实现不可否认性但不能隐藏购买者的身份信息;为了解决购买者的身份信息能够被销售商获得的问题,学者们又提出了匿名指纹方案。下面分别概述一下这些指纹协议,具体的内容分别参见 4.6、4.7 和 4.8 节。

1. 对称指纹

Benny Chor 等人在 1994 年提出了对称指纹的概念^[51],含指纹作品为用户和发行商共同拥有,所以发行商和用户都知道该作品,当发现被非法分发的带有某用户指纹的作品时,无法确定谁应对它负责,因为作品可能是该用户分发的,也可能是发行商本人分发以对该用户进行陷害。

对称协议研究的重点是使用类似纠错码或其他方法来构造唯一的“指纹”,并使用概率论的工具来证明协议的安全性。它们主要关心的安全问题是防止 U 个人的共谋攻击,这类协议中最具有代表性和可行性的是 Traitor-Tracing 协议,它不依赖其他技术的研究发展情况,主要针对广播加密节目方面,其应用性非常强。

对称性数字指纹协议的缺点:一是考虑的安全因素不够多,仅仅针对共谋攻击。二是随着共谋人数的增加,协议执行需要的数据量会大大增加,就这一点来说,如何在保证安全性的条件下降低协议的复杂性是一个重要的研究方向。这很大程度依赖于纠错码技术的发展。虽然该协议存在弊端,但提出了基本的指纹协议体系结构,奠定了数学基础,是后来指纹协议研究的基石。

2. 非对称指纹

在对称指纹方案中,当发现被非法分发的带有某用户指纹的拷贝时,将无法确定谁应该对它负责。针对这一问题,Pfitzmann 和 Schunter 引入了非对称(Asymmetric)指纹的概念^[52],以改善协议的安全性和公平性。类似于非对称的加密体制能够实现不可否认性,非对称指纹体制最主要的特点是实现非法用户的不可否认性。在此协议中,带有用户指纹的作品对发行商是不可见的,当一旦发现非法拷贝的作品,发行商可以追踪出叛逆者并能向第三方提供证据。非对称数字指纹有以下几个方面的含义^[52]:① 只要协议正常执行,用户可以得到含有其指纹的合法拷贝;② 对发行商而言,在一定的共谋尺寸下,发行商能够从非法拷贝中追踪出至少一个叛逆者,同时能够提供证明用户有罪的证据(该证据是不可伪造的);③ 对用户而言,无论共谋人数的多少,无辜用户不能受到陷害。

非对称指纹体制一般由 4 个基本协议组成:初始化协议(用户进行购买登记和发行商的有关初始化工作)、指纹添加协议(为用户生成带指纹的拷贝)、追踪协议(确认叛逆者的身份)、仲裁协议(发行商向第三方提供用户有罪的证据)。目前,非对称指纹体制的构造手段主要有基于一般的安全多方计算协议、利用特殊的密码学协议、利用密码算法等。安全多方计算是自然的实现思路,实际也是基于密码协议的研究。如 Biehl 和 Meyer 提出了一种基于常数轮零知识证明和 ANDOS 承诺协议的非对称指纹协议,但是协议执行需要较大计算量和通信量,并且需要购买者保留协议数据用于仲裁。此外,这些方案所用 Boneh-Shaw 编码长度过长是绝大多数媒体所无法达到的,且无法保证不可感知性和鲁棒性,因此不适合于感知性媒体的构造。因为对其有效实现的研究不充分,人们便寻求能够避免使用一般的安全多方计算协议的设计方法。以特殊的、更为具体的密码协议为基础和直接应用密码算法便是两种典型的思路。此外,将公钥密码算法与防篡改硬件相结合也是一种设计非对称数字指纹体制的思路。这种方法的特点是设计思想简单,但这种体制的安全性不仅基于密码算法的安全性,而且还基于硬件的安全性。尽管从实际应用角度看实现效率较高,但同时增加了硬件开销。因此,如何利用特殊的密码协议或者直接利用密码算法的有关研究成果,设计有效实用的非对称指纹体制是非常有意义的应用基础研究工作。

值得注意的是,非对称指纹协议的设计与指纹的编码是密切相关的。指纹编码所讨论的主要问题是使指纹体制能够抵抗共谋攻击。学者们已经深入讨论抗共谋攻击能力较强的非对称指纹体制的设计,但大多基于安全多方计算;而不基于一般安全多方计算的非对称指纹体制对抗共谋攻击能力的分析还需要进一步深入研究。如何设计既有较好抗共谋攻击能力,又有较好实现效率的非对称指纹方案;或在已有的具有较好抗共谋攻击能力的指纹编码方案的基础上,设计能有效实现的非对称指纹协议,都是值得进一步研究的方向。

3. 匿名指纹

无论是对称还是非对称指纹协议,用户均需在购买过程中提交自己的身份信息,这破坏了购买过程的隐秘性。正是在这种背景下,Pfitzmann 和 Waidner 于 1997 年在文献[53]中提出匿名(Anonymous)数字指纹。在该指纹机制中,用户在购买拷贝的过程中不会泄漏自己的身份信息,但如果用户进行非法分发活动,凭借非法拷贝中的信息,发行商可识别非法者的身份。匿名指纹协议的实现通常是引入一个登记中心,负责为用户的真实身份进行登记,同时为用户发放购买过程中需要的一些验证信息(通常是假名及其相应的证书)。匿名指纹的含义有两种:一种匿名的含义较弱,用户在购买带有自己指纹

的拷贝时可以对身份进行保密；但如果发行商和登记中心进行联合则可以确定出用户的身份，即这种意义上的匿名不能抵抗发行商和登记中心等实体的共谋，因此也被称为准匿名的。较强意义的匿名协议中，即使发行商和登记中心或其他实体进行联合，也不能确认无辜用户的身份，同时也不能对同一用户的不同购买进行联系。下面给出匿名指纹的主要组成及主要要求。

一个匿名指纹体制通常包括以下 4 个实体：发行商、用户、登记中心和仲裁者。由 4 个主要部分组成：登记（初始化）协议、带指纹拷贝的生成、非法用户身份确认、仲裁协议。一个匿名指纹体制应满足以下要求，① 正确性：只要协议各个部分都能够成功执行，最后用户将得到其要购买的拷贝；② 无辜用户的安全性：无辜的用户不会受到陷害；③ 发行商的安全性：当发现非法拷贝时，发行商能够凭借其中的指纹信息对用户进行追踪；④ 匿名性和不可联接性：如果没有拿到用户非法分发的拷贝，发行商（即使与登记中心进行联合）也不能确定用户的身份，即匿名性（Anonymity），而且发行商（即使与登记中心进行联合）也不能将同一个用户的不同购买进行联系，即不可联接性（Unlinkability）。

自从匿名数字指纹的概念于 1997 年被提出以来，其研究得到了许多重视。在文献[53]中，Pfitzmann 和 Waidner 给出了匿名数字指纹体制的模型，并讨论了几种变形；同时给出了在某些假设下实现匿名指纹的一种框架，但该框架过于依赖零知识证明而不具备实用价值。文献[60]提出了一种发行商一旦发现重新分发的拷贝，无需被怀疑用户及登记中心的参与，就能确定出叛逆者的匿名指纹体制。该体制的安全性基于计算离散对数的困难性及安全多方计算的可行性。在文献[61]提出的匿名指纹体制中，用户在登记中心登记获得证书，基于盲传输（Oblivious Transfer）并使用智能卡来实现指纹的嵌入，从而避免了使用安全多方计算协议。当发行商发现非法拷贝时，首先获得相应证据，在登记中心的帮助下确认用户身份。在文献[62]中，用户在登记中心登记并获得相应的证书，通过发行商与用户之间使用带有承诺的盲传输（Committed Oblivious Transfer）协议，用户得到他要购买的拷贝。但该机制存在一个很大的缺点，即对用户的追踪过程需要所有用户的参与，对无辜用户而言显然是不必要的负担。在文献[63]中，采用类似于文献[60]的匿名实现策略，但是对部分协议的实现效率做了改进。不过，文献[60~62]以及类似的匿名协议^[63]实际上并不能提供真正的匿名性，因为登记中心和发行商如果进行联合，可以将他们就一个用户的有关记录进行对比和联系，从而确定出用户的身份，因此只能称之为准匿名的。文献[64]使用代理签名构造了匿名指纹协议，但该协议只是用一个代理来减少用户的计算负担，如果协议中的登记中心和发行商进行联合，仍然可以将一个用户的不同购买进行联系，因此仍然属于准匿名的。

第一个具有严格意义上匿名性的指纹方案设计是文献[65]。在该文中，作者基于电子货币的不可追踪性，将用户的分发过程视为重复花费设计了一种匿名指纹体制。该体制使用两类比特承诺方案，一类是基于离散对数的比特承诺方案，用于对嵌入内容进行承诺并用来构造证明嵌入内容合理性的零知识证明协议；另一类基于二次剩余的比特承诺方案用于生成带指纹拷贝的步骤。后者使得带指纹拷贝的生成效率很低，因而在实际中不实用。此外在该体制中，用户需要参加仲裁协议才能最终确定其是否有罪。文献[66]在该方案的基础上进行了改进，使用户不需要参与仲裁协议，但带指纹拷贝的生成方法仍然沿袭了文献[65]中的方法。2001 年，Kuribayashi 和 Tanaka^[67]利用具有同态性质的 Okamoto-Uchiyama 加密体制^[68]提高了文献[65]、[66]中指纹嵌入协议的效率。但该文中

对匿名性的讨论并不完整, 并未真正讨论具体的登记协议, 其主要贡献在于提出了效率较高的具体的非对称指纹嵌入机制。Camenisch 在文献[69]中利用群签名构造了一种匿名指纹体制, 用户的身份是在群签名的签名群体中, 因此可以被有效隐藏。发行商可以从非法拷贝中提取出用户的秘密信息, 并行使撤销管理人 (Revocation Manager) 的职能。如果使用的群签名体制中撤销管理人能够不需成员管理人 (Membership Manager) 的帮助即可实现用户身份确认, 则发行商也可以不需登记中心的帮助直接实现对用户身份的追踪。在文献[70]提出的匿名追踪体制中, 作者在考虑匿名性时引入了登记中心、电子货币发放中心等实体, 但在该电子货币发放中心与登记中心等实体共谋的情况下, 用户的匿名性得不到保障。

由此可见, 目前对于匿名指纹体制的研究主要采用特殊的密码协议 (如群签名协议), 或者借鉴电子货币中的相关技术, 尽管已经构造了一些体制, 总体来看还很不成熟。总的来看, 在现有的匿名非对称指纹体制中, 主要有以下几个问题没有得到充分考虑: ① 大部分体制要求嵌入的内容能够被完整提取出来, 目前尚无有效的数字指纹编码方案能够实现这一点; ② 大部分带指纹拷贝的生成依赖于安全多方计算协议, 因此效率较低; 不仅如此, 使用安全多方计算的协议中, 对用户的相关承诺信息进行验证, 也需要在安全多方计算协议中进行, 以保证嵌入信息与用户承诺信息的一致性; ③ 在不使用安全多方协议的指纹体制中, 如何保证用户能够诚实嵌入自己的指纹信息 (或诚实地按照约定对该信息进行编码), 则成为研究中的难点问题。比较常用手段是使用零知识证明, 如文献[65]。

总之, 数字指纹要在实际中得到应用还需假以时日。一方面, 数字指纹协议要具备一些符合实际需求特性 (如匿名等); 而且, 随着人们版权保护意识的逐步增强, 可能还会有其他新的需求。另一方面是协议实现的效率 (目前较常用的零知识证明、比特承诺、安全多方计算等协议使得效率很低), 怎样实现既具有符合实际需求的一些特性又有较高效率的数字指纹协议, 将是数字指纹研究中的重点和难点问题。

4.5 统计指纹技术

1983 年, Wanger 提出了基于假设校验的统计指纹方案^[48]。假定有 N 个实数和 U 个用户, 假定有足够多的样本数据使得能够对统计假设进行校验, 通过对所有 u_i (用户 i) 检查 μ_i ($\mu_i = \mu_i^l - \mu_i^h$) 值就可识别是哪个用户进行了非法分发拷贝。下面介绍其详细过程。

假定 N 个实数为 s_1, s_2, \dots, s_N , U 个用户为 u_1, u_2, \dots, u_U , 假定样本数据足够多。为了让 N 个实数在统计指纹中适合使用, 必须能对每一个 s_j , 找到 δ_j ($\delta_j > 0$), 使得对每个 $i \neq j$, s_j 的 δ_j 邻域与 s_i 的 δ_i 邻域不相交。然后, 每个用户在闭区间 $[s_j - \delta_j, s_j + \delta_j]$ 中获得一个数值, 它不同于其他用户获得的数值。大致上, 一半用户获得的数值在区间 $[s_j, s_j + \delta_j]$ 上, 另一半在区间 $[s_j - \delta_j, s_j]$ 中。发送给用户 u_i 的第 j 个数据版本记做 s_{ij} 。

假定数据以某种方式被盗用, 并且发行商能从他找到的非法拷贝中提取出数值 s'_1, s'_2, \dots, s'_N 。对每个 i ($1 \leq i \leq U$), 我们想校验这个假设, 即返回数值是源自用户 u_i 的。为此, 对给定 i , 我们校验如下似然统计量

$$L_{ij} = \frac{s'_j - s_{ij}}{\delta_j}, \quad 1 \leq j \leq N \quad (4.22)$$

即, $\{L_{ij}\}_{1 \leq j \leq N}$ 是返回数值 s'_j 和给定用户 u_i 数值 s_{ij} 间的归一化差。

对于给定 i , 我们考虑两个不相交的子集 $\{L_{ij}\}_{1 \leq j \leq N}$ 的均值。令 μ_i^h 为这样一些 L_{ij} 的均值, 其 s_{ij} 是在右半区间 $[s_j, s_j + \delta_j]$ 取值的; 令 μ_i^l 是另外一些 L_{ij} 的均值, 其 s_{ij} 是在左半区间 $[s_j - \delta_j, s_j]$ 取值的, 那么可得 $\mu_i^h \leq 0$ 和 $\mu_i^l \leq 0$ 。如果令 $\mu_i = \mu_i^l - \mu_i^h$, 假定攻击者在这些值返回为 s_j' 之前不对这些数值进行修改。如果攻击者从用户 u_i 中获得数据, 那么 $\mu_i = \mu_i^l = \mu_i^h = 0$ 。如果他是从别人处获得的, 那么我们期望

$$\mu_i^h \approx -0.5, \mu_i^l \approx 0.5, \mu_i^l - \mu_i^h \approx 1 \quad (4.23)$$

因此, 如果没作修改, 除非 N 非常小, 否则应该能立刻识别出攻击者。

当攻击者改变了返回值, 甚至对较大的 N , $\mu_i^h \approx 0$, 可能也不再能识别攻击者, 这是因为攻击者可能根据一些分布, 使用非零数据修改这些值。然而, 可以假定攻击者不能区分两种可能取值中哪一个较大和哪一个较小。因此对足够大的 N , 如果攻击者的值源于用户 u_i , 可以期望 μ_i 接近于 0。另一方面, 若攻击者的值不源自用户 u_i , 对大的 N , 我们能期望

$$\mu_i = \mu_i^l - \mu_i^h \approx 1 \quad (4.24)$$

因此, 我们可用下面的算法: 对每个 i 计算出上面两个均值的差值 μ_i 。如果对某一个 i , μ_i 接近于 0, 并且对于所有其他的 $j \neq i$, μ_j 接近于 1, 那么这就为盗用数据是源自用户 u_i 的假设提供了证据。通过对所有 i 检查 μ_i 值, 就能识别出是哪个用户泄露了信息。

由于此种指纹方案是基于假设校验的, 我们可以提高假设校验的可信度。然而, 假设毕竟是假设, 不能变成确定性事实。

4.6 对称指纹技术

对称数字指纹方案假定内容提供商和消费者都是诚实可信的, 指纹拷贝的生成由内容提供商单独完成, 无需消费者的参与, 因此对称数字指纹方案实现起来简单。在这种模式中, 由于发行商和用户都知道该拷贝以及该拷贝中的指纹, 当发现被非法分发的带有某用户指纹的拷贝时, 将无法确定谁应该对它负责。因为拷贝可能是该用户分发的, 也可能是发行商本人分发的以对该用户进行陷害。下面首先介绍对称指纹的总体方案, 然后详细介绍其中的基本叛逆者追踪协议。

4.6.1 对称指纹技术的基本方案

对称指纹方案由两个算法所构成: 分发算法与追踪算法, 如图 4.3 所示。其中, 图 4.3 (a) 表示的是分发过程, 图 4.3 (b) 表示追踪过程。 m 表示用户信息, c 表示原始的数字拷贝, Fingerprinting 表示指纹过程 (指纹调制与指纹嵌入), s 表示消费者所获得的含有指纹的数字拷贝, Record 表示销售商的销售记录, s' 表示产品提供商找到的未授权数字拷贝 (盗版拷贝), Tracing 表示追踪过程。对称指纹方案的交易过程可以描述为:

- (1) 用户向发行商提出购买请求;
- (2) 发行商为用户生成相应的指纹信息 w , 并嵌入到数字产品 c 中生成含指纹拷贝 s ;
- (3) 发行商把含指纹拷贝 s 发送给用户, 并把销售记录 Record (包含 c 、用户信息 m 、相应的指纹信息 w 等) 保存到销售数据库;
- (4) 当发行商在某处发现未授权的数字拷贝 s' 时, 便启动追踪程序 Tracing 试图找到相应的叛逆者。

4.6.2 基本叛逆者追踪协议

1. 基本思想

若仅一个人知道秘密，而该秘密出现在晚间新闻上，那么其罪犯行为是显然的。然而当知道这个秘密的人数量很大时，一个更复杂的情况就出现了。如果他们所有人共享完全相同的数据，那么就解决不了谁有罪和谁无辜这个问题。从秘密共享者中找出叛逆者的一种可能方法是给所有共享者一个稍有差别的秘密。由此，Benny Chor 等于 1994 提出基本的叛逆者追踪（Traitor-tracing）协议^[51]应用这种思想来解决盗版问题。针对这个应用，需要识别出是否正在进行盗版，并防止信息传送给盗版用户且又不损害合法用户。进一步说，应该提供盗版标识的法律证据，在符合上面要求的条件下，给出一个叛逆者追踪方案。通常，称 Benny Chor 等提出的方案为 Chor-Fiat-Naor 方案。

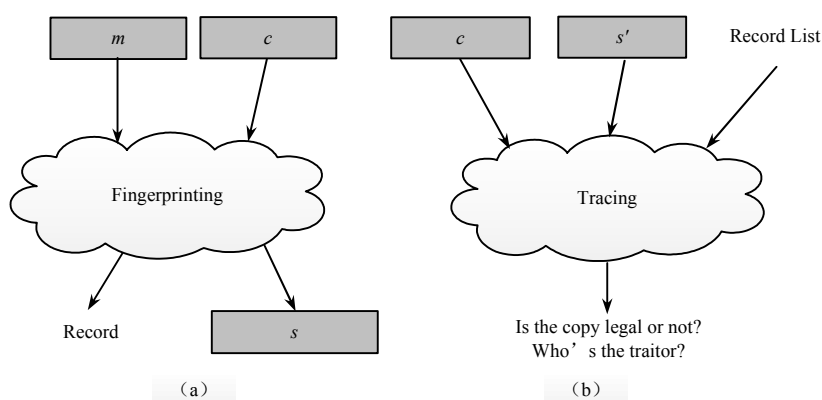


图 4.3 对称指纹技术的基本方案

为了了解 Chor-Fiat-Naor 方案，首先来了解付费电视广播系统中常用的方法。发行商向他的所有合法用户广播加密的电视节目，对每一个节目块（Block）使用不同的对称加密密钥，为了使合法用户能够解密所有节目块，发行商给每个合法用户分发一个唯一的**个人密钥**（Personal Key）。注意，为了能和不对称密码方案中的**私钥**（Private Key）加以区别，这里使用个人密钥这个词。用户使用他的个人密钥解密出节目块中的对称加密密钥，从而解密出节目块中的内容。这个个人密钥就可理解为协议中的“指纹”。

这里有几个假设：① 任何用户都需要使用一个有效的个人密钥来获取节目的内容，而不能直接把节目的明文广播给非法的用户（这个做法可能是经济代价上不可行的）；② 发行商不可能用每个合法用户的个人密钥直接加密节目或者说向每个合法用户广播完全不同的加密节目内容，因为这样广播代价太大而不可行；③ 节目必须被分成许许多多的小块，每个块使用不同的对称密钥加密（每个对称密钥本身的大小相对于节目块的大小可以忽略）。不然的话，如果节目使用同一密钥加密，则一个合法用户只要用他的个人密钥解密出对称密钥发给一个非法用户后，非法用户就可一直解密所有的节目。所以，任何叛逆者只能把自己的个人密钥发给非法用户使用，或者几个共谋叛逆者之间通过比较他们各自的个人密钥来拼出一个不同于他们的合法用户的个人密钥。当某个想发布盗版的合法用户 Alice（即叛逆者）把他的个人密钥非法告诉一个非法用户 Bob，使 Bob 也能解密出各节目块中的内容，从而构成侵权行为。当发行商发现 Bob 的侵权行为后，他可得到 Bob 所使用的个人密钥，发现该密钥的原始所有者，从而抓到盗版源头 Alice，达到保护版权的目的。

为了实现上述目的, Chor-Fiat-Naor 方案的基本思想如下: 发行商生成一个含有 r 个随机密钥的集合 X , 并从 X 中为每个用户分配 L 个密钥组成该用户的个人密钥 (是包含 L 个密钥的 X 的子集)。这里, 记第 j 个用户 u_j 的个人密钥为 PK_j 。值得注意的是, 不同用户的个人密钥可能有非空交集。一个叛逆者追踪消息由多对使能块和加密块组成, 如图 4.4 所示。每个加密块 CB_i 是在某个会话密钥 K 下对实际数据 (如几秒钟的视频剪辑) 的对称加密。每个使能块 EB_i 能够允许某个授权用户 u_j 计算得到会话密钥 K 。每个使能块 EB_i 由发行商基于部分或所有 r 个密钥而生成的加密值组成。每一个用户 u_j 通过使用他的个人密钥 PK_j 对那些他拥有密钥 (是指个人密钥 PK_j 中的那 L 个密钥) 的加密值进行解密, 从而计算出 K 。也就是说, 用户 u_j 的个人密钥 PK_j 和使能块 EB_i 对于相应的加密块 CB_i 而言是两个输入, 用来计算会话密钥 K , 如图 4.5 所示。有了会话密钥 K , 就可以解密 CB_i 得到相应的明文块 TB_i 。

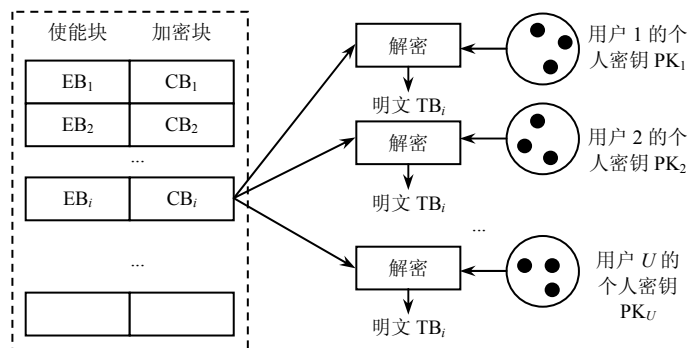
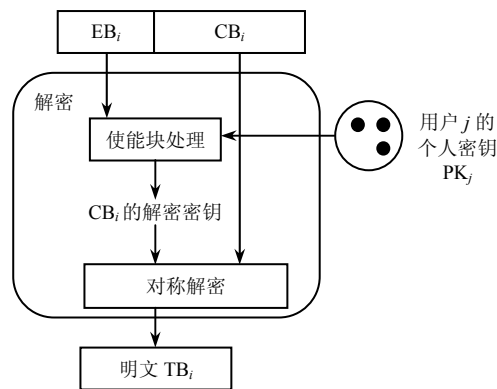


图 4.4 Chor-Fiat-Naor 方案

图 4.5 明文 TB_i 的解密过程示意图

叛逆者可能共谋, 并提供他们个人密钥的子集给某一非授权用户, 使得非授权用户也能利用他能解密出的使能块数值来计算出会话密钥 K , 然后用 K 来解密加密块。因此, 给用户分配密钥时, 需要达到以下目标: 当俘获一个盗版译码器, 并对它拥有的密钥进行检测时, 应至少能检测出一个叛逆者。为了了解叛逆者追踪协议如何保证在叛逆者共谋的情况下, 发行商也可以至少追查出一个共谋叛逆者, 或者说这个概率是很大的。下面详细介绍基本叛逆者追踪协议模型。

2. 叛逆者追踪指纹协议中的角色、对象和参数

叛逆者追踪协议中的角色包括以下几种。

- (1) **发行商 (Data Supplier 或 Merchant)**: 提供数字产品的商家。
- (2) **合法用户 (User 或 Buyer, 简称用户)**: 购买数据发布者产品的正版用户。
- (3) **叛逆者 (Traitor)**: 向非法用户提供个人密钥 (指纹) 的一个合法用户。
- (4) **共谋者 (Traitors)**: 一组通过比较各自个人密钥进行侵权活动的叛逆者。
- (5) **非法用户 (Pirate User)**: 使用叛逆者或共谋者提供的个人密钥获得发布者产品的非正版用户。

(6) **仲裁者 (Judge)**: 仲裁发行商对叛逆者或共谋者的诉讼。

叛逆者追踪协议中的对象包括以下几种。

(1) **商品 (Item)**: 数据发布者出售的一个产品 (如一部电影), 合法用户的一个个人密钥在一个同一个商品中都是有效的。每个商品被分成多个会话数据块。

(2) **会话密钥 (Session Key)**: 对称密码体制的密钥 (如流密码中的密钥)。

(3) **会话数据块 (Session Block)**: 数据发布者向外广播 (公布) 的一段产品数据, 每个会话数据块包含两部分, ① **加密块 (Cipher Block)**, 由会话密钥加密的有效数据 (明文), 每个会话数据块中使用的会话密钥都不相同; ② **使能块 (Enabling Block)**, 用分段并且对称加密的方法保存对应加密块使用的会话密钥。

(4) **个人密钥 (Personal Key)**: 每个合法用户用来解密会话密钥的一个密钥集合。

叛逆者追踪协议中的参数包括以下几种。

- (1) C : 协议可以保证安全的最大的共谋者的个数;
- (2) L : 会话密钥的分段数, 也等于个人密钥的集合大小;
- (3) B : 一个任意字母表的大小, 如字母表为 $\{1, 2, 3, \dots, B\}$;
- (4) U : 用户的最大数量。

3. 叛逆者追踪协议中的子协议

一个基本的叛逆者追踪协议可分成四个子协议, 分别介绍如下。

(1) 密钥初始化子协议

密钥初始化子协议由发行商执行。对于一个商品, 发行商随机选择 L 个对称密钥集合, 每个集合有 B 个密钥, 可形象比作一个密钥桶 (Bucket), 这 $L*B$ 个对称密钥可按字母顺序表示为 $key_{1,1}, key_{1,2}, \dots, key_{L,B}$, 如图 4.6 所示。对于这个商品的每个会话数据块, 选择一个不同的会话密钥 K 。

(2) 密钥指纹处理子协议

密钥指纹处理子协议如下: 对于每个用户 u_j , 发行商随机选择一个长 L 的码字 $word_j$, 其每个元素都是字母表 $\{1, 2, 3, \dots, B\}$ 中的某个字母。这样一个码字就和 L 个密钥桶中的 L 个密钥对应起来, 这 L 个密钥构成用户 u_j 的个人密钥 PK_j 。如码字 $word = \{7, 3, 4, \dots, 9\}$, 则个人密钥为 $PK_j = \{key_{1,7}, key_{2,3}, key_{3,4}, \dots, key_{L,9}\}$ 。发行商保存每个用户的个人密钥, 把个人密钥发给对应的用户, 个人密钥对其他人是保密的。码字的可能空间为 $B^L \gg U$ 。

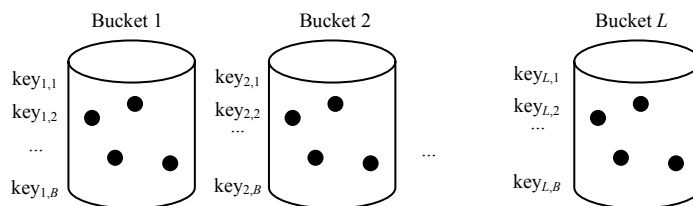


图 4.6 密钥桶示意图

(3) 会话发送子协议

会话发送子协议如下：发行商广播（发布）商品中每个会话数据块，每个会话数据块中的加密块用这个块对应的会话密钥 K 加密，会话密钥 K 被等长分成 L 段，即 $K=\{K_1, K_2, \dots, K_L\}$ ，其中第 i 段 K_i 使用第 i 个密钥桶中的 B 个密钥进行加密（ $i=1, 2, \dots, L$ ）。这 $L*B$ 个加密值按次序排列构成这个会话数据块的使能块，随加密块一起发送。每个用户 u_j 收到一个会话数据块后，使用自己的个人密钥 PK_j 先从使能块中找到自己能够解密的 L 个加密值，然后用 PK_j 中的 L 个密钥分别解密出 K_1, K_2, \dots, K_L ，组成完整的会话密钥 K 。然后，用 K 解密出对应加密块中的有效数据。

(4) 叛逆者追踪子协议

叛逆者追踪子协议如下：对一组共谋者（由 C 个叛逆者组成）来说，为了使一个非法用户可以成功解密出所有会话数据块中的数据，他们必须提供一个有效的个人密钥。首先他们不会直接提供他们当中某个人个人密钥，否则发行商直接从检测出的这个个人密钥可以查到叛逆者。因此，必须对会话密钥 K 中的每个 K_i ，从共谋者个人密钥对应的 C 个 $key_{i,l}$ （ $l=1, 2, \dots, B$ ）中选择一个作为解密 K_i 的密钥，使最终共谋产生的个人密钥是有效的而和共谋者中的任意一个个人密钥都不同，这样才能避免追查到他们中的任何一个。为防止这种共谋方式的成功，叛逆者追踪子协议通过如下算法使发行商仍能确定其中的一个叛逆者：当事后查到了一个非法用户，发行商可得到一个该非法用户使用的个人密钥 PK' （实际是每个密钥桶中对应一个 $key_{i,l}$ ）。对于 PK' 中的 L 个 $key_{i,l}$ （ $i=1, 2, \dots, L$ ），发行商对每个其个人密钥中对应位置等于 $key_{i,l}$ 的用户作一次标记。例如，假设 $PK'=\{key_{1,7}, key_{2,3}, key_{3,4}, \dots, key_{L,9}\}$ ，而某个用户的个人密钥为 $PK_j=\{key_{1,7}, key_{2,8}, key_{3,4}, \dots, key_{L,6}\}$ （设省略号位置中的对应 $key_{i,l}$ 都不相同），则这个用户 u_j 被标记两次（第 1 和第 3 两个密钥桶）。最后被标记次数最多的用户就被认为是共谋者之一。实际上，这个用户至少被标记 L/C 次，并且共谋所伪造的个人密钥等于一个合法用户的个人密钥的概率根据参数设置是可忽略的，以保证共谋不能陷害一个诚实的用户，这个算法以及共谋不可陷害的证明见文献[51]，在此不再赘述。

4.7 非对称指纹技术

在现实应用中，“发行商和用户都是诚实可信的”是一种非常理想化的假设，是很难成立的。因此，必须考虑到交易中一方或者双方都不是诚实可信的情况。对称指纹方案在现实应用中存在三个主要的问题。首先，由于指纹的生成与嵌入由发行商单方面完成，发行商知道用户所得到的最终含指纹拷贝，因此不诚实的发行商能够通过直接重新分发用户的含指纹拷贝来诬陷无辜的用户，给用户造成损害。其次，不诚实的用户能够辩称盗版的数字拷贝来自于发行商本身，从而能够抵赖自身的盗版行为。最后，当发生版权纠纷时，发行商无法向第三方仲裁机构提供相应的证据来指证叛逆者，而用户也无法提供相应的证据来证明他确实是无辜的，从而使得版权纠纷无法解决。

为克服对称指纹方案的缺陷，保障产品提供商与消费者双方的公平性，折衷办法是把指纹方案设计成非对称的。为此，Pfitzmann 和 Schunter 提出了非对称（Asymmetric）指纹的概念，率先把密码学理论引入到数字指纹技术中，并构造出第一个非对称指纹方案^[52]。类似于非对称的加密体制能够实现不可否认性，非对称指纹体制最主要的特点是实现非法用户的不可否认性。下面首先介绍非对称指纹的基本方案，然后详细介绍其中的非对称叛逆者追踪协议。

4.7.1 非对称指纹技术的基本方案

非对称指纹方案的基本方案如图 4.7 所示。图中的 text 是用于签名验证身份的一个字符串。在非对称指纹方案中，为了获得相应的数字产品，消费者必须输入一些个人的秘密信息如密钥 ($K_{S,B}$) 等。

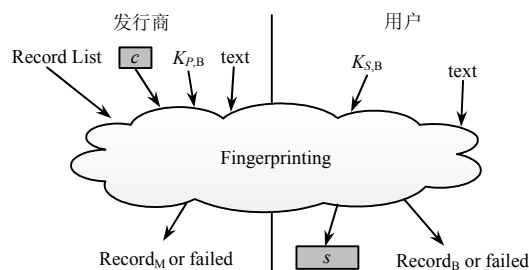


图 4.7 非对称指纹技术的指纹处理过程

非对称指纹方案的交易过程可以描述如下。

- (1) 用户向发行商提出购买请求。
- (2) 用户产生一个密钥对 ($K_{S,B}$, $K_{P,B}$)，分别为私人密钥和公开密钥，公开密钥 $K_{P,B}$ 向发行商或可信的第三方进行公开。
- (3) 通过私人密钥 $K_{S,B}$ 生成相应的指纹信息 w ，并嵌入到数字产品 c 中生成含指纹拷贝 s ，用户和发行商各自保存销售记录 $Record_B$ 和 $Record_M$ ；
- (4) 当在某处发现未授权的数字拷贝 s' 时，发行商便可启动验证程序 (identification) 确认拷贝的合法性，如图 4.8 左边所示。如出现纠纷，可通过第三方进行辩论 (disputation)，如图 4.8 右边所示。其中，proof 表示证据，acc 为仲裁结果。

非对称数字指纹有以下几个方面的含义。

- (1) 只要协议正常执行，用户就可以得到含有其指纹的合法拷贝，而发行商能够得到一些与消费者有关的一些秘密信息。

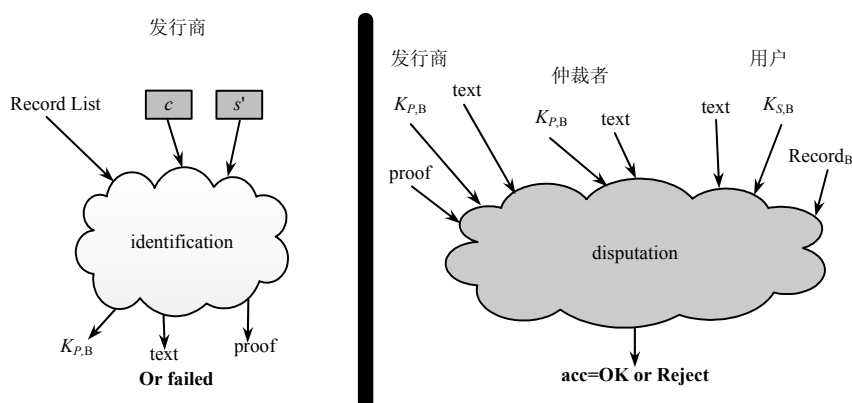


图 4.8 非对称指纹技术的验证过程和仲裁过程

- (2) 对发行商而言，在一定的共谋尺寸下，能够从盗版的数字拷贝中找到至少一个不诚实的用户 (叛逆者)，同时能够提供充分的不可伪造的证据来指证这个叛逆者。

- (3) 不可抵赖性 (Non-Repudiation)。不可抵赖性又称之为卖方 (发行商) 的安全性 (Seller's Security)，它是指不诚实的消费者不能抵赖其盗版行为。

(4) 不可诬陷性 (Non-Framing)。不可诬陷性又称为消费者的安全性 (Customer's Security)，它是指只要交易完成后，发行商不能够诬陷无辜的消费者。

(5) 对用户而言，无论共谋人数的多少，无辜用户受到错误指控的概率应该控制在一个很小的范围之内。

下面，详细介绍 Pfitzmann 和 Schunter 提出的非对称叛逆者追踪协议。

4.7.2 非对称叛逆者追踪协议

1. 基本思想

Pfitzmann 和 Schunter 将非对称的概念应用于前面 Chor 等提出的基本叛逆者追踪协议，以改善基本叛逆者追踪协议的安全性和公平性。简单来说，在基本的叛逆者追踪协议中，发行商和用户都知道卖给用户的个人密钥，那么在发现盗版提起诉讼时，用户完全可以声称发行商发现的个人密钥是发行商的一个不诚实的员工所流传出去的，从而使仲裁无法确认谁是叛逆者，这就是协议的对称性所引起的。

非对称的概念就是要使标识用户的“指纹”只有用户自己知道（自己选择生成其个人密钥的码字），而发行商在商品交易时不知道，但可以确认。在追踪时发行商可以得到足够的证据 (Proof) 来指认相关的叛逆者。非对称叛逆者追踪协议的基础仍基于前面的基本叛逆者追踪协议，包括参与的角色和对象。为了实现非对称性，需要增加三个密码学体系：一个公钥签名模式、一个承诺模式和一个安全 2-party 计算。在这里不讨论这些模式的概念和实现，假设都已存在。

2. 非对称叛逆者追踪协议中的参数

非对称叛逆者追踪协议中的参数包括如下五个方面。

- (1) σ ：一个适当的概率参数以表示安全的系数；
- (2) C ：协议可以保证安全的最大的共谋者的个数；
- (3) U ：用户的最大数量；
- (4) L ：会话密钥的分段数，等于个人密钥的集合大小，这里选 $L=64 * C * (\sigma + \log_2 U)$ ；
- (5) B ：一个任意字母表的大小，如字母表为 $\{1, 2, 3, \dots, B\}$ ，这里选 $B=48 * C$ 。

3. 非对称叛逆者追踪协议中的子协议

非对称叛逆者追踪协议分成六个子协议，分别描述如下。

(1) 用户签名密钥生成

每个用户生成一对签名模式使用的公私钥 $(K_{S,B}, K_{P,B})$ 并公布其 $K_{P,B}$ 。

(2) 密钥初始化子协议

与基本对称协议相同，发行商选择 L 个密钥桶 (Bucket)，每个桶有 B 个对称密钥。

(3) 密钥指纹处理子协议

用户选择一个长为 L 的码字 word_B (注意，这一步在基本对称协议中是由发行商完成的)。用户对 word_B 生成一个承诺 com_B ，并把这个承诺发送给发行商。用来揭示承诺的信息称为 open_B 。用户使用自己的签名私钥 $K_{S,B}$ 对消息 $\text{msg}_B = (\text{text}, \text{com}_B)$ 进行签名，生成 sig_B 。其中 text 是一个字符串，只要足以说明签名消息的含义。发行商使用 $K_{P,B}$ 验证 sig_B 是对 msg_B 的一个有效的签名。接下来执行如下安全 2-party 计算。

- 1) 输入。① 用户：码字 word_B 、用于揭示承诺的信息 open_B ；② 发行商：在密钥

初始化子协议中生成的 $L*B$ 个对称密钥、一个随机选择的大小为 $L/2$ 的集合 set_B (它是 $\{1, 2, \dots, L\}$ 的子集)、承诺 com_B 。

2) 计算。① 验证 open_B 可以打开承诺 com_B , 否则协议失败; ② 验证 set_B 的大小最多为 $L/2$ 个元素; ③ word_B 中符号对应集合 set_B 中元素的位置组成的有序符号集称为 halfword_trace_B , 其余部分称为 halfword_evid_B 。

3) 输出。① 用户: 由 word_B 对应生成的个人密钥; ② 发行商: halfword_trace_B (这样发行商只知道 word_B 中的一半符号而不能猜出整个 word_B , 仍符合非对称性)。③ 发行商得到的交易记录 $\text{record}_M = (\text{ID}_B, \text{text}, \text{com}_B, \text{sig}_B, \text{set}_B, \text{halfword_trace}_B)$, 而用户得到的交易记录 $\text{record}_B = (\text{text}, \text{word}_B, \text{open}_B)$ 。ID_B 是用户的序列号。

(4) 会话发送子协议

与基本叛逆者追踪协议中对应的子协议相同。

(5) 叛逆者追踪子协议

与基本叛逆者追踪协议中一样, 发行商发现一个非法用户, 得到其个人密钥, 也就相当于找到一个长为 L 的码字 word_F 。他搜索所有交易记录中的 (set_B , halfword_trace_B), 找到一个记录, 其 halfword_trace_B 和 word_F 至少有 $L/(4*C)$ 个对应位置上符号是相同的。他取出 $\text{msg}_B = (\text{text}, \text{com}_B)$ 和 sig_B 准备向仲裁者控告此用户。

(6) 仲裁子协议

发行商组成的证据为 $\text{proof} = (\text{msg}_B, \text{sig}_B, \text{word}_A)$, 其中 word_A 是长为 L 的码字, 码字中对应 set_B 中元素的位置上由 halfword_trace_B 中的对应符号组成, 码字中其他位置的符号由 word_F 中的对应位置符号组成。此外, 用户提供 open_B 。仲裁者首先验证 sig_B 签名的有效性和承诺的有效性 (揭示 word_B)。然后验证 word_A 和 word_B 是否至少在 $L/2 + L/(16*C)$ 个对应位置上的符号相同, 成立则指控成功; 否则用户即可否认指控。为了使 word_B 始终对发行商保密, 可以用零知识证明来给出 word_B 。

4.8 匿名指纹技术

无论是对称还是非对称指纹协议中, 用户均需在购买过程中提交自己的身份信息, 这破坏了购买过程的隐秘性。正是在这种背景下, Pfizmann 和 Waidner 在文献[53]中提出了匿名数字指纹 (Anonymous Fingerprinting) 的概念。这种指纹机制中, 用户在购买拷贝的过程中不会泄漏自己的身份信息, 但如果用户进行非法分发活动, 凭借非法拷贝中的信息, 发行商可以识别非法者的身份。匿名指纹协议的实现通常是引入一个可信第三方 TTP (Trusted Third Part), 负责为用户的真实身份进行登记, 同时为用户发放购买过程中需要的一些验证信息 (通常是假名及其相应的证书)。下面先介绍匿名指纹技术的基本思想, 然后介绍一种典型的匿名指纹实施方案。

4.8.1 匿名指纹技术的基本思想

匿名指纹技术实际上和其他匿名技术是密切相关的, 必须有相应的匿名技术作为前提, 才能保证用户的私人信息不被泄露。1983 年, Chaum 介绍了一种签名方案^[71], 它能让一个签名者在数据上签名, 而不泄漏数据的内容, 也就是所谓的盲签名。匿名指纹是盲签名的一个应用, Pfizmann 和 Waidner 介绍了基于一个可信任第三方的匿名非对称指纹和非对称指纹方案^[53]。买方能以匿名的形式购买信息。但是如果他们非法重新发布这

种信息, 仍然能被识别出来, 具体流程参照图 4.9。电子市场应该与传统市场一样, 提供相似的隐私权。也就是说, 在实际的电子购买中应该获得一定的匿名性。当人们用现金买商品时, 在传统市场上没有人能追踪这次购买。仅为了使用户用指纹标识其个人本身, 而破坏所有这样的匿名性是不值得的。

匿名指纹的基本思想如下: 买方选择一个假名 (即签名方案中的一个密钥对 $(K_{S,B}, K_{P,B})$), 然后用他的真实身份对其签名, 表示他对这个假名负责。他从注册中心获得一个证书 cert_B , 有了这个证书, 注册中心宣布它知道选择这个假名的买方的身份 (也就是, 注册中心能把一个真人用一个假名代表)。然后当买方进行一次购买时, 在不了解商人的情况下用标识这次购买的文本 text 计算出一个签名: $\text{sig}=\text{sign}(K_{S,B}, \text{text})$, 然后将信息 $\text{emb}=(\text{text}, \text{sig}, K_{P,B}, \text{cert}_B)$ 嵌入购买的数据中。他在一个比特承诺里隐藏这个值, 并以零知识方式给商人发送证书和承诺。比特承诺是一种密码技术, 它能传递数据并仍能在一段时间内保持秘密, 当需要鉴别时, 商人提取出 emb 并给注册中心发送 $\text{proof}=(\text{text}, \text{sig}, K_{P,B})$, 并要求验证。作为回答, 注册中心向商人返回用户的签名。于是, 商人能够使用这个签名来验证所有的值并提供证据 proof 指控买方。

4.8.2 一种典型的匿名指纹技术

这里介绍一种由孙中伟和冯登国提出的一种基于同态公钥加密体制的匿名数字指纹方案^[72]。该方案不仅保证了含指纹数字媒体对发行商是不可见的, 同时隐匿了用户身份。由于利用公钥加密算法的同态性质, 数字指纹方案在保证非对称嵌入的同时, 实现相对简单。

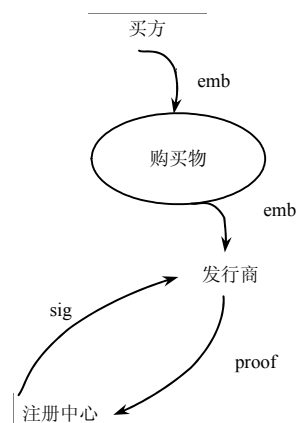


图 4.9 匿名指纹技术的基本思路

1. 数字指纹嵌入机制

数字指纹的嵌入可以通过数字水印嵌入算法来完成, 它既可以加性嵌入方式嵌入到原始媒体数据中, 又可以乘性嵌入方式嵌入到原始媒体数据中。在原始媒体数据的时空域或变换域, 如果数字指纹以乘性方式嵌入到原始媒体数据之中, 嵌入规则为

$$s_i = (1 + \alpha w_i) c_i, \quad i = 1, 2, \dots, N \quad (4.25)$$

而以加性方式嵌入到原始媒体数据之中, 则嵌入规则为

$$s_i = c_i + \alpha w_i, \quad i = 1, 2, \dots, N \quad (4.26)$$

其中 $C=\{c_1, c_2, \dots, c_N\}$ 为原始载体序列, $S=\{s_1, s_2, \dots, s_N\}$ 是嵌入指纹后的序列, $W=\{w_1, w_2, \dots, w_N\}$ 为嵌入的指纹信号, α 为指纹嵌入的强度因子。

对于定义在代数结构 $(G, *)$ 上的公钥加密函数 $E: G \rightarrow R$, 如果给定 $E(x)$ 和 $E(y)$, 其中 $x, y \in G$, 在没有私钥的情况下能够计算出 $E(x * y)$, 则称该公钥加密系统具有同态性质。例如, RSA 公钥密码算法满足乘同态性, 即 $E(x) \cdot E(y) = E(x \cdot y)$ 。而 Paillier 公钥密码体制满足加同态性, 即 $E(x) + E(y) = E(x + y)$ 。同态公钥密码体制目前在安全投票协议、多方计算和签名等方面得到了广泛的应用。

若数字指纹采用乘性方式嵌入到原始的媒体数据中, 并采用 RSA 公钥密码体制对其进行加密, 由 (4.25) 式可知

$$E(s_i) = E(1 + \alpha w_i) \cdot E(c_i), \quad i = 1, 2, \dots, N \quad (4.27)$$

若数字指纹采用加性方式嵌入到原始的媒体数据中, 且采用 Paillier 公钥密码体制对其进行加密, 由 (4.26) 式可知

$$E(s_i) = E(c_i) + E(\alpha w_i), \quad i = 1, 2, \dots, N \quad (4.28)$$

2. 匿名指纹方案

协议的参与实体有: 发行商 (D)、用户 (B)、注册中心 (R)、仲裁者 (A)。其中注册中心为可信的第三方。基本协议有: 指纹生成协议、指纹嵌入协议、跟踪协议、仲裁协议。我们以数字指纹采用乘性嵌入方式为例来说明孙中伟和冯登国构造的匿名指纹方案。首先给出建立匿名指纹方案的系统条件: ① 假设存在一个鲁棒的数字指纹检测与提取算法 E_x 。② 发行商 D、用户 B 和注册机构 R 使用 RSA 公钥密码体制来完成数字媒体的加密及解密操作, 并设嵌入强度 α 为 1。其中加密算法为 $E(\cdot)$, 解密算法为 $D(\cdot)$ 。③ 发行商 D、用户 B 和注册机构 R 各自拥有经过认证的 RSA 公钥/私钥对, 分别为 (PK_D, SK_D) 、 (PK_B, SK_B) 和 (PK_R, SK_R) 。另外, 他们还拥有经过认证的签名函数 sign 和验证函数 ver , 以及与之对应的密钥对。

(1) 指纹生成协议

指纹生成协议是在用户 B 向发行商 D 和注册中心 R 提出购买请求之后, 在用户 B 和注册中心 R 之间执行的双方协议, 过程如下。

1) 用户 B 选择自己的匿名公钥/私钥对 $(K_{P,B}, K_{S,B})$, 并生成自己拥有匿名私钥 $K_{S,B}$ 的证据 proof , B 对匿名公钥 $K_{P,B}$ 和 proof 签名, 将签名结果 $\text{sign}_B(K_{P,B}, \text{proof})$ 、 proof 以及 $K_{P,B}$ 发送给注册中心 R, 并请求生成一个数字指纹。

2) 注册中心 R 首先验证用户 B 的签名 $\text{sign}_B(K_{P,B}, \text{proof})$, 如果验证通过, 则 R 为用户 B 随机地生成数字指纹 $w = \{w_1, w_2, \dots, w_N\}$; 否则, 协议终止。

3) 考虑式 (4.27) 以及 $\alpha=1$, 注册中心 R 用 $K_{P,B}$ 加密待嵌入的指纹, 记

$$E_{K_{P,B}}(w) = \{E_{K_{P,B}}(1 + w_1), E_{K_{P,B}}(1 + w_2), \dots, E_{K_{P,B}}(1 + w_N)\} \quad (4.29)$$

并将 $K_{P,B}$ 和 $E_{K_{P,B}}(w)$ 以及对它们的签名 sig 发送给 B, 其中 sig 计算如下

$$\text{sig} = \text{sign}_R(E_{K_{P,B}}(w), K_{P,B}) \quad (4.30)$$

4) 用户 B 验证注册中心 R 关于签名 sig 的真实性。若通过验证, 则解密 $E_{K_{P,B}}(w)$ 获得指纹 w ; 否则, 协议终止。

(2) 指纹嵌入协议

指纹生成后, 接下来可以执行嵌入协议, 过程如下:

1) 用户 B 将 $K_{P,B}$ 、 $E_{K_{P,B}}(w)$ 以及注册中心的对它们的签名 sig 发送给发行商 D。

2) 发行商 D 验证 sig 的真实性。如果验证通过, 则继续下一步; 否则, 协议终止。

3) 发行商 D 产生一个用来唯一标示与该用户交易的指纹 v , 并将它嵌入到用户 B 将要购买的原始媒体数据 c 中, 得到 $c' = c + v$ 。

4) 发行商 D 产生一个随机置乱 π 作用于用户 B 发送过来的 $E_{K_{P,B}}(w)$ 上, 计算结果为

$$\pi(E_{K_{P,B}}(w)) = E_{K_{P,B}}(\pi(w)) \quad (4.31)$$

5) 发行商 D 将 $\pi(w)$ 作为第二个数字指纹嵌入到 c' 中。注意, 尽管 w 已经用用户 B 的公钥 $K_{P,B}$ 进行了加密, D 得到的只是 $E_{K_{P,B}}(w)$, 然而利用 RSA 公钥密码体制的同态性, D 在不解密出 w 的情况下可按下面的计算方法完成 $\pi(w)$ 的嵌入

$$E_{K_{P,B}}(c'') = E_{K_{P,B}}(c') \cdot \pi(E_{K_{P,B}}(w)) = E_{K_{P,B}}(c' \cdot \pi(w)) \quad (4.32)$$

6) 发行商 D 将 $E_{K_{P,B}}(c'')$ 发送给用户 B, 同时将 v 、 $K_{P,B}$ 、 $E_{K_{P,B}}(w)$ 、 sig 和 π 作为发

行记录保存起来。

7) 用户 B 用自己的私钥 $K_{S,B}$ 解密 $E_{K_{P,B}}(c'')$, 得到加了数字指纹的媒体数据, 即

$$s = D_{K_{S,B}}(E_{K_{P,B}}(c'')) = c' \cdot \pi(w) = (c + v) \cdot \pi(w) \quad (4.33)$$

(3) 跟踪协议

发行商 D 一旦发现盗版拷贝 s' , 便利用相应的指纹提取和解码算法提取标示与用户交易的指纹 v' 。若提取失败, 协议终止; 否则, D 在其发行记录中找到与指纹 v' 最相似的 v , 并找到与 v 关联的用户信息 $K_{P,B}$ 、 $E_{K_{P,B}}(w)$ 、 sig 和 π 作为起诉非法分发用户的证据。

(4) 仲裁协议

仲裁协议由发行商 D 提出, 它是发行商 D、用户 B 以及仲裁者 A 三方执行的协议, 目的是让仲裁者 A 证明用户 B 是数字媒体的非法分发者。

1) 发行商 D 将盗版证据 v 、 $K_{P,B}$ 、 $E_{K_{P,B}}(w)$ 、 sig 和 π 提交给仲裁者 A, A 首先验证注册中心 R 的签名 sig 。如果验证通过, 则继续下一步; 否则, 协议终止。

2) 仲裁者 A 要求用户 B 出示其数字指纹 w , 并用 $K_{P,B}$ 加密 w 。若加密结果与 $E_{K_{P,B}}(w)$ 不符, 那么 B 提交的是一个伪造的指纹, 则认为 B 是非法用户; 如果符合, 则继续下一步。

3) 仲裁者 A 根据 π 和 w 计算 $\pi(w)$, 并检测 $\pi(w)$ 是否存在于盗版拷贝 s' 中, 若存在, 则认为用户 B 是叛逆者; 否则, 则认为 B 是无辜的。

3. 安全性分析

一个匿名指纹系统对购买数字产品的用户来讲, 应该满足其购买行为的匿名和不可关联, 并且发行商无法诬陷用户。而对发行商来讲, 当用户在非法分发了其购买的数字产品之后, 发行商能根据非法分发的数字产品追踪到该用户, 并且用户无法否认。同时, 由于匿名指纹系统涉及多个参与实体, 它还须具备共谋容忍的特点。

(1) 匿名及不可关联性

用户 B 除了具有由证书认证中心 R 颁发的公钥、私钥对 (PK_B, SK_B) 之外, 对于每次购买行为, 他还生成了匿名的公钥、私钥对 $(K_{P,B}, K_{S,B})$, 而注册中心是可信的第三方, 不会与发行商进行共谋, 因此, 用户购买行为的匿名性能够得到保证。另外, 只要用户购买某个数字多媒体作品, 匿名指纹协议就会为该次购买行为产生一对匿名的公钥、私钥。因此, 无法通过两个数字作品来判断购买的是否属于同一个人, 也就满足不可关联的要求。

(2) 发行商的安全性

如果所采用的公钥加密算法和签名体制是安全的, 那么恶意的用户无法修改或替换由注册中心产生的数字指纹。发行商 D 为了维护自己的利益, 在数字作品中嵌入了 v 和 $\pi(w)$, 尽管用户知道数字指纹 w , 但他不能移去 $\pi(w)$, 因为不知道发行商对 w 进行的置乱 $\pi(\cdot)$ 。如果发行商 D 发现了非法分发的数字作品, 那么只要存在一个鲁棒的指纹提取算法, 他就可以追踪到非法分发用户 B。而用户 B 也无法对自己的行为进行反驳, 因为指纹嵌入是在加密的状态下进行, 只有用户 B 能解密并获得带指纹的数字作品 s 。另外, 使用时间戳能够阻止 B 用以前的指纹替代当前的指纹。

(3) 用户的安全性

为了伪造一个带特定指纹的数字多媒体作品, 发行商 D 要么知道用户 B 的私钥 $K_{S,B}$, 用以解密 $E_{K_{P,B}}(w)$, 而私钥 $K_{S,B}$ 只有用户 B 知道; 要么 D 直接得到 B 的指纹 w ,

这必须通过注册中心 R 方可实现,但是, R 是可信的第三方,因此发行商 D 不能与 R 共谋得到 B 的指纹 w 。因此,对用户 B 来说,发行商 D 没法诬陷用户 B 。

4.9 共谋安全指纹技术

盗版者为了避免被抓到,必然要去破坏作品中的数字指纹信息,故数字指纹方案在设计时就应考虑盗版者的攻击。抵抗滤波、压缩、变形等攻击的方法在数字水印技术研究中已经提出很多,这些方法同样可应用到数字指纹中,这里不作深入讨论。但是,由于嵌入在同一个作品中的各个用户的指纹信息不同,数字指纹还面临着来自多个盗版者的共谋攻击,因此抗共谋攻击是数字指纹研究的一项独特的主要内容。我们通常把能够抵抗共谋攻击的指纹技术称为共谋安全指纹技术。目前的共谋安全数字指纹方案可分为两种:一种是基于**嵌入假设**的离散型数字指纹,它在编码的过程中包含了离散编码的步骤,该指纹设计方案认为不可探测比特是不可改变的,或者只容忍很少的改变,其代表是 C -安全码。另一种是未采用嵌入假设的连续型数字指纹,这种指纹的每个分量取值范围是实数,而不像离散型数字指纹那样每个分量只能取有限种状态,正交码是连续型数字指纹的代表。离散型数字指纹方案的基础是嵌入假设,但是嵌入假设并不是总是成立的,某些攻击将会破坏嵌入假设。连续型抗共谋指纹的每一位在连续域上取值,在追踪盗版者的时候,将从盗版作品中提取的指纹和每个用户的指纹进行相关性检测,根据相关度来判断某个用户是否参与了共谋,由于不是基于嵌入假设,所以对不可探测位置的修改不很敏感,不必受嵌入假设的限制。离散型指纹编码方法已经提出得比较多,而连续指纹编码方法的研究主要集中在正交码、 n -simplex 码和 OFFO 码等上。这里我们关注离散型共谋安全数字指纹技术,首先介绍共谋攻击的主要方式,接着介绍与指纹编码有关的一些术语和定义,然后介绍编码设计问题,然后介绍三种典型共谋安全编码方法,最后介绍一种典型的叛逆者追踪实现方案。

4.9.1 共谋攻击方式

由于同一数字作品的每份拷贝都不相同,带来的问题是,多个用户可以联合他们的拷贝,比较拷贝中的不同之处,从而可以找出拷贝中的部分指纹标记的位置,对这些位置上的标记进行修改,选择其中的一种甚至创造新的标记,从而制造出一份新的作品拷贝,将其分发,希望能够逃避跟踪。这一过程就称为用户共谋,是盗版用户针对数字指纹系统的主要攻击方式。共谋攻击仅针对同一作品不同拷贝中不一致的内容进行修改,是专门针对指纹标记的攻击方式。对于这类攻击,数字指纹技术主要通过数字指纹编码过程中使指纹具有抗共谋能力来对指纹加以保护。

用户共谋时,可以选择不同的共谋方式,但总的来说,共谋用户会采用某一被认为最优的策略,使他们能够成功制造出某个不属于任何用户或者属于无罪用户的指纹,从而免于责罚。通常,共谋用户的最优策略是使他们被抓获的可能性最小又或是成功陷害某个无罪用户的概率最大。如果盗版用户可以伪造多个共谋指纹时,他们将选择最符合上述策略的指纹。用户共谋产生新的作品拷贝,这一行为等同于制造新的用户指纹,数学上,可以将共谋过程理解为由用户指纹通过共谋运算得到共谋指纹。具体的共谋运算可以不仅仅是简单意义上的代数运算,可以是逻辑数学中的逻辑运算,统计数学中的统计运算,也可以是综合这些运算的多次组合运算。

下面,介绍几种常见的共谋攻击方案。

1. 逻辑与

即共谋运算选择逻辑运算中的“与”运算,共谋指纹由所有参加运算的指纹码字按位相“与”而得。在“逻辑与”的情况下,码“0”与任意码元相“与”结果都为“0”;码“1”仅在和码“1”相“与”时才仍保持为“1”。例如,用户指纹码字(0010111)、(0101101)和(0001011)以“逻辑与”方案进行共谋后,共谋指纹的码字为(0000001),生成码字的每一比特都是由共谋用户对应位置上比特值相“与”而得,即在用户指纹不同的位置上,共谋运算选择了“0”码元。

2. 逻辑或

即共谋运算选择逻辑运算中的“或”运算,共谋指纹由所有参加运算的指纹按位相“或”而得。在“逻辑或”的情况下,码“1”与任意码元相“或”,结果都为“1”;码“0”仅在和码“0”相“或”时才仍保持为0。例如,用户指纹码字(0010111)、(0101101)和(0001011)以“逻辑或”方案进行共谋后,共谋指纹的码字为(0111111),生成码字的每一比特都是由共谋用户对应位置上比特值相“或”而得,即在用户指纹不同的位置上,共谋运算选择了“1”码元。

3. 简单平均

即共谋运算为算术运算或统计运算中的求取平均值。这也是一种比较简单、共谋用户也比较容易实现的共谋策略。例如四个不同用户码字在某位置上的比特值分别为1、1、1、0,则在进行简单平均后,得到的值为 $3/4$,四舍五入为1。由此可见,平均值代表该位置上选择1的概率,如果某个标记位置上的码元“1”越多,则其均值越接近于“1”;反之,则越靠近“0”。

4. 加权平均

如果参与共谋平均的用户,在平均之前对不同用户的指纹加权,亦即共谋运算为算术运算或统计运算中加权平均,为不造成数据较大的改变,他们权值的和为1。这样会削弱一些用户指纹对共谋指纹的影响,降低他们被跟踪到的可能性;但同时会加强另一些用户对指纹的影响,使他们被跟踪到的可能性增加。这一共谋方式与所假设的用户共谋策略相抵触,但仍有可能被采用。

5. 随机选取

上述“逻辑与”和“逻辑或”的策略可看作两种极端情况,更多的情况是“随机选取”策略,也就是共谋用户比较他们的拷贝,在发现不相同的位置上,随机地选择两种形式中的一种,选择其中任一形式的概率为50%。广义来看,共谋策略都可看作是在可探测位置上,以某一概率 p 选择两种码元中的一种,以 $1-p$ 的概率选择另一种。“随机选取”中的 $p=0.5$,而“逻辑与”中选择码元“0”的概率为1,“逻辑或”中选择码元“1”的概率为1。

6. 最大策略

在这一策略中,共谋用户拷贝中不相同的位置上,选择共谋集中该位置上出现次数较多的码元;若某位置上各码元出现次数相同,则任选其一。这一策略相对复杂,但较为符合共谋策略假设。当共谋用户数为2时,这一策略退化为“随机选取”策略。例如,用户指纹码字(0010111)、(0101101)和(0001011)以“最大策略”方案进行共谋后,共

谋指纹的码字为(0001111), 位置 2 上“0”码元出现了 2 次而“1”码元仅出现 1 次, 共谋集中该位置上“0”码元出现次数较多, 故而“最大策略”选择码元“0”作为共谋指纹在该位置上的比特值。

7. 最小策略

这一策略与“最大策略”相反, 共谋用户在可探测位置上, 选择共谋集中该位置出现次数较少的码元; 若某一可探测位置码元出现次数相同, 则任选其一。当共谋用户数为 2 时, 同样退化为“随机选取”策略。例如, 用户指纹码字(0010111)、(0101101)和(0001011)以“最小策略”方案进行共谋后, 共谋指纹的码字为(0110001), 位置 2 上“0”码元出现了 2 次而“1”码元仅出现 1 次, 共谋集中该位置上“1”码元出现次数较少, 故而“最小策略”选择码元“1”作为共谋指纹在该位置上的比特值。

4.9.2 术语与定义

为了研究方便, 在讨论指纹编码时, 不关心数字作品的形式, 而仅讨论指纹码字, 将指纹等同于含指纹数据。同样, 不关心标记形式和标记的具体位置而只讨论指纹码比特和码比特的位置。通常假定指纹由二进制比特序列组成, 有 $\{0, 1\}$ 两种可能的标记状态。叛逆者进行共谋时, 如果拷贝中处于某个位置上的标记有所不同, 他们就能探测到该标记。相反, 如果处于某个位置上的标记都相同, 他们将无法探测到该标记。如果用户改变不可探测标记的状态, 将导致作品无法使用, 用于指纹标记的数据必须具备这一特性。共谋用户能够将可探测标记改为 0、1 两种状态之一, 甚至使其成为不可读状态?, 即无法确定该标记是 0 还是 1。因此, 共谋用户能将可探测标记改为 $\{0, 1, ?\}$ 三种状态之一而无法修改不可探测标记。数字指纹编码技术中假设指纹数据满足上述性质, 是指纹编码的一个重要的前提假设, 称为**嵌入假设**。需要指出, 指纹被嵌入到数据后, 盗版用户无法区分可探测标记值中的 0 值和 1 值, 共谋盗版所能找的只是可探测标记的位置。

假设每个含指纹数据中嵌入了一个长度为 N 的二进制码向量 $w \in \{0, 1\}^N$, 即指纹每个比特对应于数据中嵌入的一个标记。用户用 $1, 2, \dots, U$ 标号。嵌入到用户 u_i 作品中的指纹用 w_i 表示, $i \in \{1, 2, \dots, U\}$ 。所有用户指纹码字的集合表示为 $\Gamma = \{w_1, w_2, \dots, w_U\}$, 每个码的码长为 N , 集合大小为 U 。可以看出, 如果指纹码长为 N , 就可能有 2^N 种指纹, 但只有 $w_i \in \Gamma$ 的指纹是用户指纹码集中的码字。

定义 4.1: 给定 N 维向量 $w \in \{0, 1\}^N$ 和集合 $I = \{i_1, i_2, \dots, i_r\} \subseteq \{1, 2, \dots, N\}$, 其中 $r \leq N$, 用 $w|_I$ 表示向量 $(w|_{i_1}, w|_{i_2}, \dots, w|_{i_r})$, 其中 $w|_{i_j}$ 表示 w 的第 i_j 位 ($j=1, 2, \dots, r$), 称向量 $w|_I$ 为 w 在位置集合 I 上的限定。

例如, 用户指纹 $w=(001101)$, 若 $I=\{1, 3, 5, 7\}$, 那么 w 在位置 I 上的限定为 $w|_I=(0011)$ 。

定义 4.2: 如果 $G=\{w_{t_1}, w_{t_2}, \dots, w_{t_r}\} \subseteq \Gamma$ 为排序后的用户指纹码子集, t_1, t_2, \dots, t_r 表示 r 个用户的标号 ($r \leq N$), 定义 G 的第 j 个输入向量为 $w^{(j)}=(w_{t_1|j}, w_{t_2|j}, \dots, w_{t_r|j})$, 即向量 $w^{(j)}$ 顺序包含了指纹码集 G 中每个指纹的第 j 个比特。

例如, 两用户指纹为 $w_1=(0001101)$ 和 $w_2=(0010001)$, 顺序观察指纹 w_1 和 w_2 , 可得 $G=\{w_1, w_2\}$ 的 7 个输入向量分别为 $w^{(1)}=(w_1|_1, w_2|_1)=(00)$, $w^{(2)}=(00)$, $w^{(3)}=(01)$, $w^{(4)}=(10)$,

$w^{(5)}=(10)$, $w^{(6)}=(00)$, $w^{(7)}=(11)$ 。

定义 4.3: 如果 $G \subseteq \Gamma$ 的某比特位置上对应的输入向量的各元素都相等, 称 G 的该比特位置为不可探测的。

例如, 两用户指纹为 $w_1=(0001101)$ 和 $w_2=(0010001)$, 可以看出 $G=\{w_1, w_2\}$ 在比特位置 1, 2, 6, 7 是不可探测的, 因为这些位置上对应的输入向量中的元素全为 0 或全为 1。实际上, 某个比特位置不可探测等价于满足嵌入假设时用户共谋不可能探测到该位置。

定义 4.4: 设 $G \subseteq \Gamma$ 为一个用户共谋集, I 为 G 的不可探测比特位置集, 则 G 的可用集定义如下:

$$F(G, \Gamma) = \{w \in \{0, 1\}^N, w|_I = w_l \mid \text{for } \forall w_l \in G\} \quad (4.34)$$

其中, l 为 G 中的任意某个用户的标号。

例如, 设两用户指纹 $w_1=(0001101)$ 和 $w_2=(0010001)$ 构成 $G=\{w_1, w_2\}$, $I=\{1, 2, 6, 7\}$ 为用户 1 和用户 2 的不可探测位置的集合。根据定义 4.4, 知道这组盗版用户的可用集为 $F(G, \Gamma) = \{00 \times \times \times 01\}$, \times 表示该码元任意为 0 或者为 1。可见, 可用集包含所有与共谋用户的不可探测比特位置相匹配的指纹。换句话说, 共谋用户集 G 中所有码字在 G 的某个不可探测位置上的值都相同, 且可用集 $F(G, \Gamma)$ 中所有码字在该位置上的取值也都与这个值相同; 对于共谋集的所有不可探测位置, 可用集都满足这一性质; 集合 $F(G, \Gamma)$ 由所有满足该条件的码字组成。可见, 共谋集 G 包含于共谋可用集 $F(G, \Gamma)$ 中。一般情况下, 我们只考虑某一种指纹编码的情况, 可用 $F(G)$ 表示 $F(G, \Gamma)$, 隐含所涉及的指纹码集 Γ 。

用可用集的定义, 可规范地陈述前文提及的嵌入假设如下。

定义 4.5 (嵌入假设): 共谋只能产生包含于共谋可用集 $F(G)$ 中的某个指纹的对象。如果用户将某个可探测标记改为 “?” (不可读), 可强制使它置 0 或置 1。由此, 可定义嵌入假设下的共谋如下。

定义 4.6: 设 $G=\{w_1, w_2, \dots, w_g\}$ 为一个共谋集, 码长为 N , 集合大小为 g , 共谋产生共谋指纹 z , $F(G)$ 为 G 的可用集。共谋策略 $f(\cdot)$ 满足: $z=f(G)$ 且 $z \in F(G)$ 。

该定义建立了共谋数学模型, 将其看作由共谋用户指纹码字参与运算的函数, 用户指纹通过函数运算得到一个共谋码字。根据嵌入假设, 共谋码字必定落在共谋用户指纹的可用集中, 并且在共谋用户指纹集中, 实际参与共谋运算的只有共谋集的可探测比特。

4.9.3 编码设计

指纹编码是数字指纹技术的核心内容, 也是数字指纹技术中最关键的部分。指纹编码要充分考虑生成指纹码字的抗共谋能力。指纹编码的模型和一般的编码原理模型类似。把指纹码集视作码长为 N , 大小为 U 的二进制码, 将指纹问题看作一个使用信道编码或者纠错控制编码的通信问题。通信中一个或多个码字被选定用于传输, 传输的码字对应于指纹问题中共谋用户的指纹。传输过程并非理想传输, 对应于非法指纹的接收码字是由传输码字以不为人知的某种方式“混合”而成, 即用户进行了共谋。在接收端, 发行商用接收到的码字对传输码字(集)进行估计。在通信中, 这一情形被称为“多址接入”, 用接收码字判定所发出的码字称为“译码”。

在共谋攻击的追踪过程中，在“逻辑与”共谋方案下，根据所获得的共谋指纹，发行商无法断定指纹中的“0”比特是来自不可探测位置还是由用户共谋得到，因此他只能通过指纹中的“1”比特来对共谋用户进行追踪。“逻辑或”方案则正好相反，发行商只能通过其中的“0”比特来进行追踪。对于“简单平均”和“随机选取”共谋情况，发行商对于比特“0”和比特“1”都无法确定它们是来自不可探测位置还是由用户共谋得到，因此无法完成对共谋用户的追踪。

通过采用码字扩展技术，即将码字的码位进行扩展来实现对共谋方式的追踪，发行商可以以较大的概率获得共谋用户的不可探测位置。例如，用序列“1...10...0”对码字“1”进行扩展，用其补码“0...01...1”对码字“0”进行扩展，则可以同时实现对“逻辑与”和“逻辑或”追踪的要求。在共谋攻击中，共谋仍无法改变不可探测位置上的码比特，即不可探测位置上码比特“0”和“1”的形式不发生改变；而可探测位置上的码比特几乎全部有所改动。因此检测时，如果检测到某个码比特扩展位置上的序列能够严格构成扩展序列“1...10...0”或“0...01...1”时，可判断该码比特为不可探测码比特；反之，若不能构成扩展序列，则判断该码比特为可探测码比特，用“?”表示。若码串中的“0”、“1”个数均为 d ，共谋策略中以 p 的概率选择某一码元，以 $1-p$ 的概率选择另一码元，则他们成功将某一可探测位置码比特改为不可探测位置码比特的概率为

$$P_f(d) = \frac{2}{p^d(1-p)^d} \quad (4.35)$$

当 d 较大时，这一概率无疑是很小的，但这实际上相当于整体增加了码字长度。

对“简单平均”攻击方式，我们在码字提取时通过提高对码字“1”或“0”的检测门限来进行检测判决，再按照“逻辑与”的方式进行追踪。对于“随机选取”攻击方式，由于共谋用户选取“1”或“0”的概率各为 0.5，因此我们可以通过加长码字扩展序列来满足追踪的要求。假设扩展序列为“101010”，则共谋用户将某码字所有扩展位的位置上的码字选择为“1”的概率为 $1/64$ ，当检测到某码字扩展位置上的码字不能构成扩展序列时，我们判断其为“0”，然后按“逻辑与”的方式对共谋用户进行追踪。

对于“最大策略”和“最小策略”，码字扩展技术失效。需要指出，码字长度也是指纹编码的一个重要参数，不同的编码方案长度要求也不同。我们要求码字的利用率高，而且在同样的保护能力下，码字长度要尽可能的短。下面考察一些实际的编码方案。

4.9.4 C-安全码

C-安全码是 Boneh 和 Shaw 提出的一种编码方案^[50]，全称为对数长度 C-安全编码方案 (Logarithmic Length C-secure Code)，也有人称 B&S 码。前面 4.4.1 小节已经做了一些说明，这里再进行一些补充说明。

首先，需要给出汉明距离和码字重量的定义如下。

定义 4.7: 码字 $w_1 \in \{0,1\}^N$ 和码字 $w_2 \in \{0,1\}^N$ 的汉明距离 $d_h(w_1, w_2)$ 定义为对应比特分量不相同的数目。例如 $w_1=(0001101)$ 和 $w_2=(0010001)$ 的汉明距离为 $d_h=3$ 。

定义 4.8: 码字 $w \in \{0,1\}^N$ 的重量 wt 定义为 w 中码元“1”的数目，显然 $wt(w)=d_h(w, \mathbf{0})$ ，其中 $\mathbf{0}$ 表示所有码元为 0 的 N 维向量。

带有错误概率 ε 的 n -安全的 (l, n) 码的码本 $\Gamma_0(n, d)$ 的参数 n 和 d 分别决定了码字的个数和最小汉明距离，码字的长度为 $l=(n-1) \cdot d$ 。C-安全码的码本示例 $\Gamma_0(4, 3)$ 如式

(4.21) 所示。 \mathbb{C} -安全码的码字包含有 $(n-1)$ 组的 d 位全0或者全1的标记 e_j ($j=1, 2, \dots, n-1$)。一般来说, 用户 u_i ($i=1, 2, \dots, n$) 码字 a_i 由 $(i-1) \cdot d$ 个“0”码与紧随其后的 $(n-i) \cdot d$ 个“1”码构成。这种码在整体的结构设计上比较平衡, 类似于三角矩阵。

在对共谋码字进行追踪时, 根据检测到的第一组非全0码的标记 e_l 和最后一组非全1码的标记 e_s 来判定共谋用户。由于共谋用户得不到组号小于 l 和大于 s 的嵌入位置信息, 因此无法修改这些位置上的嵌入信息。例如, 根据式(4.21)可知, 若用户1和用户3进行共谋时, 他们无法修改第三组码而会修改第一组和第二组的码, 以此来判定用户1和用户3参与了共谋。

从 \mathbb{C} -安全码的设计上, 可以看到码字长度随着用户数呈线性增加, 这样的共谋安全指纹在绝大多数实际应用场合是不能接受的。

4.9.5 I 码

I 码可以看作对标准单位矩阵取补后进行码字扩展, 主要针对“逻辑与”共谋。I 码的码本 $\Gamma_1(n, d)$ 的参数 n 和 $2d$ 分别决定了码字的个数和最小汉明距离, 码长为 $l=n \cdot d$ 。当 $d=1$ 时, 码字的长度为 $l=n$ 。一般来说, 容纳用户数为 n 的I 码 $\Gamma_1(n, 1)$, 用户 u_i 的码字在第 i 个位置上为0码, 而其他位置上皆为1码, 故可以用维数为 n 的单位矩阵取补得到。

显而易见, 当不同组合的 $g \leq n$ 个用户码字进行“逻辑与”后, 所得到的组合是唯一的, 即组合中码比特“1”的位置是唯一的。由于在“逻辑与”共谋下, 共谋码字中的比特“1”对应于共谋集中全部的不可探测位置, 因而可以唯一确定地对共谋用户追踪, 即用户数为 n 的I 码能够抵抗最大共谋用户数目为 n 的“与”共谋。采用码字扩展技术后, I 码将能够抵抗“逻辑与”、“逻辑或”、“平均”、“随机选取”等方式的用户共谋。

I 码存在的问题与 \mathbb{C} 安全码相同, 码字长度随着用户数线性增加, 同样在实际应用中是不可接受的。

4.9.6 BIBD 码

指纹编码可以看作是在一组位置中, 选取在其中的某些位置上置“1”, 其余的置“0”。因而, 可以用组合设计(Combinatorial Design)来设计指纹编码。组合学是数学的一个分支, 它用来按一定的规则选择和安排事物。组合设计则是从一给定的集合中选定一组子集以满足某种特定的性质。例如, 对子集中元素的数目或每个元素出现的次数做某些限制等。下面介绍区组设计以及BIBD设计的相关知识。

设 $\Gamma=\{t_1, t_2, \dots, t_v\}$ 为一个有限集, $B=\{B_1, B_2, \dots, B_b\}$ 是 Γ 的子集的集合, 其中 $B_i \subseteq \Gamma$, $i=1, 2, \dots, b$, 则 $\{\Gamma, B\}$ 称为 Γ 上的一个设计。 Γ 中元素的总数称为设计 $\{\Gamma, B\}$ 的阶, 记为 $|\Gamma|$ 。按照组合设计的传统定义方法, 集合 $\{\Gamma, B\}$ 中的元素被称为区组(Block)。如果至少有一个区组中没完全包括 Γ 中的所有 v 个元素, 则称该设计是不完全的(incomplete)。如果在一设计中, 各区组的容量(即区组中所包含的元素数)相同, 且 Γ 中所有各元素在 B 中出现的次数相同, 则称为该设计为一个区组设计(Block Design)。当任意一对元素 $t_i, t_j \in \Gamma$, $i \neq j$, 在 B 中相遇的次数 λ 也相同, 则称该设计是均衡的(Balanced)。

定义 4.9: 设 Γ 为包括 v 个不同元素的基集, B 为 Γ 的 k -子集(集合元素个数为 k)的集合, r 为含有某任意元素的 k -子集数, 且对任意一对元素 $t_i, t_j \in \Gamma$, $i, j=1, 2, \dots, v$, $i \neq j$, 有 λ 个区组同时包括它们, 则称 $\{\Gamma, B\}$ 构成的区组设计为均衡不完全区组设计

(Balanced Incomplete Block Design), 简记为 BIBD(v, b, r, k, λ)或 BI(v, b, r, k, λ)设计。

例如, 对于如下区组设计

$$\begin{array}{llll} B_1 & 1 & 4 & 7 \\ B_2 & 2 & 5 & 8 \\ B_3 & 3 & 6 & 9 \end{array} \quad \begin{array}{llll} B_4 & 1 & 5 & 9 \\ B_5 & 2 & 6 & 7 \\ B_6 & 3 & 4 & 8 \end{array} \quad \begin{array}{llll} B_7 & 1 & 6 & 8 \\ B_8 & 2 & 4 & 9 \\ B_9 & 3 & 5 & 7 \end{array} \quad \begin{array}{llll} B_{10} & 1 & 2 & 3 \\ B_{11} & 4 & 5 & 6 \\ B_{12} & 7 & 8 & 9 \end{array} \quad (4.36)$$

元素数 $v=9$, 区组数 $b=12$, 每个元素出现次数 $r=4$, 每个区组包含的元素数 $k=3$, 任意一对元素相遇的次数 $\lambda=1$ 。根据定义, 该区组设计可表示为 BIBD(9,12,4,3,1)。

可以证明, BIBD 存在的必要条件是

$$\begin{cases} bk = rv \\ r(k-1) = \lambda(v-1) \end{cases} \quad (4.37)$$

从该公式可知, 如已知 v, b, r, k, λ 这 5 个参数中的任意 3 个, 则其他两个参数也就随着确定了。故 BIBD(v, b, r, k, λ)也常简写为 BIBD(v, k, λ)。上例中的 BIBD(9,12,4,3,1)的参数间, 显然满足 $12 \times 3 = 4 \times 9$ 及 $4 \times (3-1) = 1 \times (9-1)$ 两个必要条件。式 (4.37) 只是 BIBD(v, b, r, k, λ)存在的必要条件, 而非充分条件。例如, 当 $v=b=22, r=k=7, \lambda=2$ 时, 式 (4.37) 中的两个条件均满足。但是, 这种参数的设计是不存在的。

下面我们再来看一下关联矩阵的定义。

定义 4.10: 设 $\Gamma = \{t_1, t_2, \dots, t_v\}$ 为一个有限集, $B = \{B_1, B_2, \dots, B_b\}$ 是 Γ 的子集的集合, $\{\Gamma, B\}$ 的 BIBD(v, b, r, k, λ)设计的关联矩阵 (Incidence Matrix) M 是一个如下的 $b \times v$ 的二进制矩阵(4-8)。

$$\begin{aligned} M &= [m_{ij}], 1 \leq i \leq b, 1 \leq j \leq v \\ m_{ij} &= \begin{cases} 1 & \text{若 } t_j \in B_i \\ 0 & \text{若 } t_j \notin B_i \end{cases} \end{aligned} \quad (4.38)$$

例如, 下面的 BIBD(7, 3, 1)矩阵 B 的关联矩阵可以表示为矩阵 M 的形式。

$$B = \begin{bmatrix} 1 & 2 & 4 \\ 1 & 3 & 6 \\ 1 & 5 & 7 \\ 2 & 3 & 5 \\ 2 & 6 & 7 \\ 3 & 4 & 7 \\ 4 & 5 & 6 \end{bmatrix} \quad M = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \quad (4.39)$$

可以证明, 对于 $\lambda=1$ 的 BIBD($v, k, 1$)设计, 对其关联矩阵 M 每个元素进行取反运算得到的码就是 $k-1$ 安全的抗“逻辑与”共谋码。例如, 按上面的 BIBD(7,3,1)生成的码矩阵 A

$$A = \bar{M} = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (4.40)$$

我们将多个用户按“逻辑与”共谋后，产生的码字 z 中的“1”称为特征码。不难看出，特征码是在共谋集的不可探测位置上标记为 1 的码。因此，可以根据共谋指纹 z 中的特征码对共谋用户进行追踪。例如，将上式 (4.40) 所示的 A 矩阵任意两个行向量进行“逻辑与”运算后，得到的特征码字都是不相同的。

使用 $BIBD(v, k, 1)$ 设计生成 $BIBD$ 抗共谋码 A 用于数字指纹抗共谋攻击时，设计中的参数 b 对应于方案可容纳的用户数 n ，参数 v 为用户指纹的码长 l ， $k-1$ 为方案允许的最大共谋用户数 C ，即在共谋用户数 g 小于或等于 $C=k-1$ 时，可追踪到所有参与共谋的用户。可以证明， $n=(l^2-l)/(C^2+C)$ 。因而，若该 $BIBD$ 存在， n 用户的指纹只需要 $l=O(C\sqrt{n})$ 个基向量，即指纹长度近似地随着 \sqrt{n} 而线性增加、随着 C 线性增加。相对于 C -安全码和 I 码， $BIBD$ 码的码字长度更加逼近理论上的下界。

因此， $BIBD$ 码能够极大地缩短指纹码字的长度。例如，存在参数为 (61, 305, 20, 4, 1) 的 $BIBD$ ，用户数为 $n=305$ ， C 安全码和 I 码需要 305 个二进制码比特，而 $BIBD$ 码则仅用 61 个二进制码比特将能成功追踪到参与共谋用户数小于等于 3 的共谋用户。在实际应用中，可认为大多数用户都是可以信赖的，因而往往不需要有非常大的允许共谋最大用户数。而在用户数目很多的情况下，码字长度往往显得非常重要。此外，在不扩展码字的情况下， C 安全码和 I 码的最小汉明距分别为 1 和 2，而 $BIBD$ 码的码距相对要大一些，即 $BIBD$ 码的抗干扰能力比 C 安全码和 I 码要强。对 $BIBD$ 码的码字进行扩展，可以使其能在“逻辑与”、“逻辑或”、“平均”和“随机选取”共谋方式下追踪共谋用户。

然后， $BIBD$ 码用作指纹也存在问题，即某参数下 $BIBD$ 设计本身的存在性及相应的 $BIBD$ 区组的获取都存在问题，尤其指纹要求参数 $\lambda=1$ ；在参数比较大的情况下，寻找 $BIBD$ 区组的算法是相当复杂的。其次，对于 $BIBD$ 共谋集的探测是通过特征码字进行追踪的，因而追踪过程的匹配算法不是将共谋指纹与用户指纹进行直接匹配检测，而是将共谋指纹与所有共谋用户数为 $g \leq k-1$ 的共谋集的特征码字进行匹配，这样的共谋用户集数量 L 为

$$L = \sum_{i=1}^{k-1} C_n^i = \frac{n!}{1!(n-1)!} + \frac{n!}{2!(n-2)!} + \dots + \frac{n!}{(k-1)!(n-k+1)!} \quad (4.41)$$

当用户数 n 或者 k 稍大时， L 就大得相当惊人了。在实际应用中，需要对 L 个用户共谋集的特征码进行存储；在追踪检测时，需要完成 L 次的匹配运算，开销是难以接受的。

4.9.7 典型叛逆者追踪方案

根据上面的介绍，这里给出基于 Boneh 和 Shaw 的 C -安全码的共谋安全指纹技术。

1. 码构造

令 Ω 是大小为 b 的字符表，表示了 b 个不同的标记状态。 Ω 中的字母可以是 1 到 b 的整数。对于二进制情况，一般采用 $\Omega=\{0, 1\}$ 。给定一个长度为 l 的码字 $z \in \Omega^l$ 和一个索引集合 $I=\{i_1, i_2, \dots, i_r\} \subseteq \{1, 2, \dots, l\}$ ，用 $z|_I$ 表示向量 $(z_{i_1}, z_{i_2}, \dots, z_{i_r})$ ，这里 z_i 是 z 的第 i 个字母， $z|_I$ 为 z 在位置集合 I 上的限定。称 $\Gamma=\{z^{(1)}, z^{(2)}, \dots, z^{(n)}\} \subseteq \Omega^l$ 为一个 (l, n) 码。对 $1 \leq i \leq n$ 中每个 i ，将码字 $z^{(i)}$ 分配给用户 u_i 。令 G 是一个共谋用户集，就是一个参与制作非法拷贝的用户集合。如果分配给 G 中所有用户的码字在第 i 个位置相匹配 ($i \in \{1, 2, \dots, l\}$)，就说位置 i 对 G 来说是不可探测的；也就是对 $G=\{u_1, u_2, \dots, u_g\}$ ， $z_i^{(1)} = z_i^{(2)} = \dots = z_i^{(g)}$ 。假定第 i 个标记对 G 来说是可检测的，那么基于 G 能产生一个指纹，该指纹的这个标记

处于一个不可读的状态，以至于检查人员不能确定不可读标记处于哪一个状态。标记一个不可读状态的标记为“?”。令 I 是 G 中所有不可探测位置的集合，那么就 G 中的任意用户 u_j 而言， G 的可行集定义为：

$$F(G) = \{z \in (\Omega \cup \{?\})^l : z|_I = z^{(j)}|_I, \forall z^{(j)} \in G\} \quad (4.42)$$

因此，可行集 $F(G)$ 包含与 G 的不可探测比特相匹配的所有码字。

基于上述定义，用下面的特性来构造码：没有一种共谋能串通起来陷害一个不在共谋集 G 中的用户。当限定 C 个用户组成的共谋集时，称这样的码为 C -防陷害码。设想一个发行商用码 Γ 来标识各个用户，并且基于用户的一个共谋集 G 共谋产生了码字 z 标识的非法用户，然后发布这个新用户。当发行商发现该非法用户时，想构造一个追踪算法来检测 G 的子集。如果有一个追踪算法 Tr 满足下面条件，那么我们就说码 Γ 是 C -安全码，即如果一个或至多 C 个用户的联合 G 生成一个码字 z ，满足 $\text{Tr}(z) \in G$ ；也就是追踪算法 Tr 针对输入 z 必须输出共谋集 G 中的一个成员。因此，根据一个非法拷贝，至少可以追踪到共谋集中的一个成员。

假定 C 个用户的一个共谋集 G 生成数字产品的一个非法拷贝，能使我们至少以 $1-\varepsilon$ 的概率捕获共谋集 G 中的一个成员，这种指纹码称之为具有误差 ε 的 C -安全码。通常，如果至多 C 个用户的一个共谋集 G 生成一个码字 z ，满足

$$\Pr[\text{Tr}(z) \in G] > 1 - \varepsilon \quad (4.43)$$

那么 Γ 是具有 ε 误差的 C -安全码。在构造共谋安全码字中，我们能获得一个追踪算法，假定某个共谋集 G 生成一个码字 z ，这个算法以概率 $1-\varepsilon$ 输出 G 中的一个成员。

令 e_i 是一高度为 n 的比特列，前 i 个比特是 1 而其余的是 0。码本 $\Gamma_0(n, d)$ 由 e_1, e_2, \dots, e_{n-1} 的所有列组成，每一个重复 d 次。重复次数决定了错误概率 ε 。令 $z^{(1)}, z^{(2)}, \dots, z^{(n)}$ 表示 $\Gamma_0(n, d)$ 的码字。例如，针对 4 个用户的码本 $\Gamma_0(4, 3)$ 见式 (4.21)。发行商在数字产品中嵌入 $\Gamma_0(n, d)$ 的码字之前，他需作如下的随机选择：随机挑选一个置换 $\pi \in Q_l$ ，这里 Q_l 是 l 个字母所有置换的完全对称群。使用码字 $\pi z^{(i)}$ 对用户 u_i 的拷贝加入指纹。注意所有的用户都应用同样的置换，并且 π 需要保密。

2. 追踪算法

令 B_i 是所有具有如下特性的比特位置集：用户在其位置能看到类型为 e_i 的列。也就是说， B_i 是如下比特位置的集合：前 i 个用户看到一个 1 并且其余用户看到一个 0。 B_i 中的元素个数是 d 。对于 $2 \leq j \leq n-1$ ，定义 $R_j = B_{j-1} \cup B_j$ 。对一个二进制串 z ，令 $\text{wt}(z)$ 表示 z 的重量 (z 中 1 的个数)。

假定用户 u_j 不是生成码字 z 的联合 G 的成员。保密的置换 π 防止共谋者知道哪一个标记代表了码本 $\Gamma_0(n, d)$ 中的哪一个比特，共谋者所知道的唯一信息是它能检测的标记值。对于不含用户 u_j 的共谋集 G ，对于一个 $i \in R_j$ 的比特位置，共谋集 G 不能区分 i 是在 B_j 还是在 B_{j-1} 中。这意味着无论他们使用哪种策略设置在 $z|_{R_j}$ 的比特，在 $z|_{R_j}$ 中的 1 以很大概率均匀地分散在 $z|_{R_j}$ 和 $z|_{R_{j-1}}$ 之间。因此，如果在 $z|_{R_j}$ 中的 1 不是以很大概率均匀分布，那么用户 u_j 是生成 z 的共谋集 G 的一个成员。通过一些计算，我们能获得在 $\Gamma_0(n, d)$ 中 d 的一种度量，对 $n \geq 3$ ， $d = 2n^2 \log(2n/\varepsilon)$ ，并且追踪算法能明确表示如下：

给定共谋码字 $z \in \{0, 1\}^l$ ，用户数 n ，找到生成 z 的共谋集 G 的一个子集的步骤如下。

(1) 如果 $\text{wt}(z|_{B_1}) > 0$ ，则输出用户 1 为参与共谋者。

(2) 如果 $\text{wt}(z|_{B_{n-1}}) < d$ ，则输出用户 n 为参与共谋者。

(3) 对于其他用户, 即 $j = 2, 3, \dots, n-1$, 令 $k = \text{wt}(z|_{R_j})$, 如果

$$\text{wt}(z|_{B_{j-1}}) < \frac{k}{2} - \sqrt{\frac{k}{2} \log \frac{2n}{\varepsilon}} \quad (4.44)$$

则输出用户 j 为参与共谋者。

需要指出, 该方法建立在**嵌入假设**的基础上, 给出了一种指纹码字长度 l 与用户数目 n 的对数及共谋容忍尺寸 C 的四次方成正比的指纹编码方案。因此, 当数据的发行量较大时, 该方法并不实用。

4.10 本章小结

本章给出了数字指纹技术的提出背景、基本概念与分类、常见的数字指纹技术及实现方法, 并重点介绍了对共谋安全的数字指纹技术。要使数字指纹技术成为真正实用的版权保护工具, 还需要进一步的研究和完善。这主要体现在以下几方面。

(1) 理论基础的研究。数字指纹技术属于信息隐藏技术中的一种。然而, 信息隐藏技术发展到今天, 还没有找到自己的理论依据, 没有形成理论体系, 很多信息理论的经典理论在处理信息隐藏系统时不再适用。随着数字指纹技术的不断发展, 它对理论指导的期待已经越来越迫切。

(2) 编码方案的研究。目前国内外数字指纹技术研究领域, 数字指纹的研究工作主要集中于数字指纹的指纹编码工作上。人们不断地努力寻找更加有效的指纹编码方案。

(3) 嵌入技术提取的完善。数字指纹系统是一种目标明确的被攻击对象, 在稳健性上需要较高的要求, 指纹嵌入提取技术还需要进一步完善, 提高其稳健性和可靠性。另外, 对于不同数据类型的嵌入提取问题和嵌入技术与压缩技术之间的关系问题也是研究中所必须考虑的。

(4) 合谋策略的分析及抗合谋研究。目前合谋攻击的研究只是针对较为常用的合谋策略, 通常都比较简单, 较复杂的合谋策略及相应的抗合谋的问题还有待深入研究。考虑同时对多种合谋形式进行跟踪, 将博弈理论用于指纹研究会是一个很好的选择。

(5) 系统整体设计研究。数字指纹技术在实用化中所需的指纹检测体系的建立、技术衡量标准、法律的保护等问题有待解决。

(6) 其他相关技术的研究。考虑将其他信息隐藏技术应用于指纹技术中, 或将指纹技术与其他技术相结合, 也可以成为指纹技术的研究方向。



习题

1. 完善的数字指纹方案应满足什么样的基本要求?
2. 数字指纹技术与数字水印技术的联系和区别分别是什么?
3. 数字指纹的攻击方式有哪些?
4. 衡量数字指纹系统的性能评价指标有哪些?
5. 常见的数字指纹协议有几种? 分别是如何实现?
6. 数字指纹系统的合谋攻击方式有哪些? 都有什么特点?
7. 请查阅有关连续指纹编码的 n -simplex 码方案, 并介绍其基本原理。

8. 参照式 (4.21) 请写出 \mathbb{C} 安全码 $\Gamma_0(5, 4)$ 的码本。设按顺序将 a_i 分配给第 i 个用户 u_i , 假设有两个用户进行了共谋, 发行商从共谋拷贝中提取的共谋指纹为 “0000101011111111”, 请问是哪两位用户进行了共谋?

9. 请给出一种符合定义 4.9 的 BIBD 设计, 并给出对应的 BIBD 抗共谋攻击码。

10. 设某数字指纹系统基于 BIBD(7,3,1) 生成的码矩阵 A 如式 (4.40) 所示。假设 7 个用户为 $\{u_1, u_2, \dots, u_7\}$, 有两个用户进行了 “逻辑与” 共谋, 发行商从共谋拷贝中提取的共谋指纹为 “0110000”。试问哪两个用户进行了共谋?

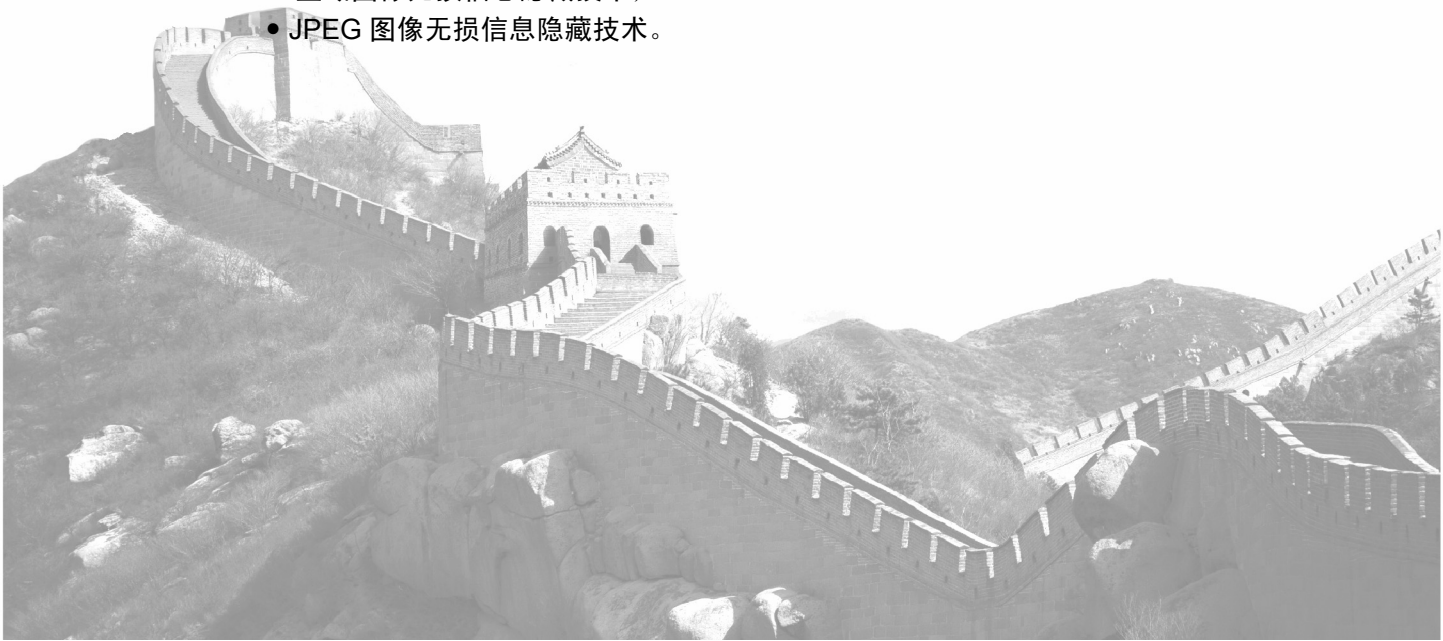
无损信息隐藏技术

本章引言

前面介绍的传统信息隐藏技术（数字水印技术和隐写术）在将秘密信息嵌入到载体对象中时，往往给载体对象带来不可恢复的永久性失真（即使提取出秘密信息后）。然而，在医学、遥感、工程制图、军事图像和法庭证据等领域，传统信息隐藏技术带来的载体对象永久性失真阻碍了传统信息隐藏技术的应用。因此迫切需要一种既能有效地避免破坏原始载体信息，又能通过隐藏信息发挥作用的新方法——无损信息隐藏（Lossless Information Hiding）方法。本章从无损信息隐藏技术的提出背景入手，首先介绍无损信息隐藏技术的相关概念、分类、框架模型和性能评价，接着以图像为载体对象介绍基于不同嵌入域（空域、整数变换域和压缩域）的无损信息隐藏技术。

本章重点

- 无损信息隐藏技术的相关概念和分类；
- 无损信息隐藏系统模型和性能评价；
- 空域图像无损信息隐藏技术；
- JPEG 图像无损信息隐藏技术。



5.1 无损信息隐藏技术的提出背景

传统的信息隐藏方法把秘密信息嵌入到载体对象的过程中往往给载体对象造成永久性失真，这成为它应用到一些对精度要求较高的领域的障碍，如法律认证、军事通信、医学图像、遥感图像和指纹图像等领域，因为这些领域既要保护待隐藏信息的安全性，又要保持载体对象本身的质量。例如，医学图像是医生对患者的生理疾病信息获取及诊断病情的一个重要依据，往往不允许有任何改动。因为改动后一旦出现误诊，很容易引起法律上的纠纷。另外，由于医学图像的获取往往代价相当高，临床上普通的一次 CT、MRI、PET（正电子发射成像）的检查都要不菲的费用，这些设备的成本都十分高昂，这与普通的数码相机获取的数字图像所需的代价形成迥然对比。不难看出，无论从法律上还是从经济成本上考虑，任何可能对医学图像造成永久性损失的操作都是不可取的。又比如，二维 CAD 工程图形广泛应用于建筑、机械、服装等领域，正规的工程图形往往具有严格的工艺约束（如形位公差、尺寸公差等），这就要求面向工程图的数字水印技术对图形的修改必须限制在工程图工艺约束的容忍范围之内，这就给研究带来了挑战。为了解决这些问题，无损信息隐藏技术就开始成为学者们研究的热点。

无损信息隐藏技术将重要信息隐藏到相对不重要的信息中，使重要信息用不易察觉或监测的方法传输，提高了信息传输的隐蔽性，从而使信息获得多一层保护；在接收端，提取出隐藏的重要信息后，完全无损恢复原载体对象，从而不影响原始载体的使用。无损信息隐藏过程一般由密钥来控制，即通过嵌入算法将待隐藏信息隐藏于公开载体中，而伪装对象（隐藏有秘密信息的公开载体）则通过公用信道传输；在接收端，利用密钥从伪装对象中提取嵌入信息和恢复原始载体对象。载体对象通常为图像、音频和视频等多媒体信息；待隐藏的信息可以是文本、图像、声音、视频等多媒体信息。随着因特网、无线通信网、数字通信和多媒体信息技术的不断发展，基于图像及视频的无损信息隐藏技术应用到公共通信领域有着较广阔的应用前景和重要的研究价值。研究基于图像及视频的无损信息隐藏方法对军事、法律、医学、遥感等领域的信息安全具有重要的意义。

5.2 无损信息隐藏技术的相关概念和分类

5.2.1 无损信息隐藏的相关概念

无损信息隐藏（Lossless Information Hiding）也称**无损数据隐藏**、**可逆数据隐藏**（Reversible Data hiding）或**可逆水印技术**，主要指在嵌入信息时，虽然可能会对载体对象的感知质量造成一定的破坏，但如果伪装对象（隐含信息的载体）在传输过程没有发生变化，那么某些合法用户和权威机构就可以根据其中所隐含的隐藏信息，清除失真后恢复原始载体。由于这种方法的独特性质，可以应用到对原始载体要求较高的很多领域，例如：在商业上，可应用于防止数字化的票据被篡改；在医学上，可以用于检测数字医疗图像传输存储的过程中是否被修改；在军事上，可以用于保证所获得的数字化军事卫星图像的完整性；在法庭上，可以检验作为证据的数码照片的真实性。

无损信息隐藏技术与一般的信息隐藏技术几乎没有原理性的差别，但是两者关注的焦点却发生了转移，传统信息隐藏技术主要关注秘密信息，而无损信息隐藏技术更注重

原始的载体信息。无损信息隐藏技术在多媒体信息传送之前,将秘密信息嵌入其中,然后再进行传输,接收端提取出嵌入信息之后,可以完全修复由于嵌入对载体信息的破坏,得到原始载体信息。无损信息隐藏算法就是可以修复嵌入过程对载体对象损坏的信息隐藏方法,其不仅传送了秘密信息,还可以得到原始载体的准确拷贝。这样一来,对于嵌入隐藏信息的要求也变得严格,主要表现在以下三个方面。

- (1) 需要知道添加数据的顺序与位置;
- (2) 需要了解原始载体数据被更改的值大小;
- (3) 避免因数据信息的嵌入而超过原始载体中的数据范围。

归结起来,无损信息隐藏技术具有以下典型特性要求。

(1) 不可检测性:指伪装对象与原始载体对象具有一致的特性,如具有一致的统计噪声分布等,使得非法拦截者无法判断是否有隐藏信息。

(2) 透明性(不可感知性):利用人类视觉系统或人类听觉系统特性,经过一系列隐藏处理,使伪装对象相对于载体对象没有明显的降低质量现象,而隐藏的数据却无人能为地看见或听见。

(3) 安全性:隐藏算法有较强的抗攻击能力,即能承受一定程度的人为攻击或传输中的信息丢失和噪声影响等,而使隐藏的信息不会被破坏。

(4) 自恢复性:提取隐藏的信息后,能精确地恢复原载体对象。

除了充分考虑以上无损信息隐藏技术的几点要求之外,通常还需考虑算法的嵌入容量和时间复杂度等方面。

5.2.2 无损信息隐藏的关键问题

1. 高嵌入容量与无损恢复

近年来的无损信息隐藏方法已在医学图像、遥感图像和军事图像等方面有了很大的进展。但是,目前的无损隐藏方法大多都是对图像经过预处理后,把最不重要的载体信息位经过无损压缩后(如用算术编码等方法进行压缩),腾出空间来嵌入待隐藏信息。这样就存在以下三个问题:① 原始载体最不重要信息经无损压缩后的输出信息量越大,待隐藏信息的可嵌入容量越小;② 在无损信息隐藏的实际应用中,嵌入容量越大,其对图像产生的失真越大,即嵌入容量与嵌入信息后的载体感官质量相互制约,当待隐藏信息容量变化时,不能自适应地选取较低失真的可嵌入区域进行嵌入;③ 当前的无损隐藏方法都是针对数字图像来考虑的,而如何做好大数据量医学动态影像信息或 MPEG-2 彩色视频等多媒体信息的管理、安全存储和传输成为了新的热点。因此,如何提高嵌入容量是无损信息隐藏系统中的一个关键问题,仅仅依靠无损压缩技术已经无法满足当前大数据量秘密信息的隐秘传输。这时,可以从可逆变换等处理方法来进行研究,尽量少的嵌入附加信息(如原始载体的修改位置信息)以及提高秘密信息的嵌入率。

无损恢复是无损信息隐藏系统中的另一个关键问题,即在接收端,信息被提取出来后,如何精确地恢复原始载体。目前大多数无损隐藏的方法都是在接收端提取出信息后,用无损编解码的方法来恢复原始载体。这样存在的问题是当无损压缩后的数据流中某段数据信息出现丢失或受噪声影响时,整个无损编码后的数据流都无法恢复出来。针对这一问题,可考虑对图像或视频帧进行分块处理,信息丢失或噪声影响只仅仅对当前块有影响,其他图像块能精确地恢复出来。

2. 无损信息隐藏的技术难点

无损信息隐藏系统主要技术难点如下。

(1) 如何在数字媒体中实现较大容量的无损隐秘传输

文本、图像等秘密信息，本身具有较大的数据量。要将大数据量的信息嵌入到另外一幅图像、一段医学动态影像或一段经过标准编码压缩过的视频数据流中，则需要较大的信息嵌入容量，而较大的嵌入容量常需对载体对象做较多的修改，这样不仅影响了嵌入信息后伪装对象的视觉效果，而且易于被察觉和被检测。因此，需要研究出具有较大嵌入容量和较高隐蔽性、安全性的无损隐藏方法。

(2) 如何在数字多媒体中实现较低失真的无损信息隐藏

秘密信息，如文本、图像等信息嵌入到另外一幅图像或视频中时，会造成原图像或视频的降质。伪装对象在传输过程中容易被察觉和被检测，如何实现嵌入信息后的伪装对象仅有较小的降质，具有较好的视觉质量或听觉质量，是一个技术难点。

(3) 嵌入算法的复杂度和实时性问题

秘密信息在发送端的嵌入过程与在接收端的提取过程，常要求能够实时完成。因此，通常要求在设计无损信息隐藏算法时，需要考虑算法的时间复杂度和空间复杂度。

(4) 嵌入率、嵌入效率及嵌入算法的安全性问题

为了提高无损隐藏算法的可嵌入容量，需要考虑算法的嵌入率和嵌入效率（嵌入率定义为单位载体可嵌入的数据量；嵌入效率定义为修改 1 比特载体对象时，可嵌入的数据量）。另外在信息传输过程中，如何保证嵌入算法的安全性也是一个重要问题。

(5) 秘密信息在传输过程中的出现信息丢失或噪声影响的问题

藏有信息的伪装对象在公共信道，如局域网、宽带网等信道上进行传输时，有可能会出现问题丢失或噪声影响的问题。如何保证所秘密信息的完全提取，以及原始载体的精确恢复，也是系统设计的一个关键问题。

5.2.3 无损信息隐藏的分类

无损信息隐藏技术可以按照不同的方式进行分类，可以按照载体对象类型分类，也可以按照嵌入域分类，还可以根据鲁棒性分类。下面作简要说明。

按照载体类型分类，常见的无损信息隐藏技术可以分为基于文本、图像、音频、矢量图和视频等各种不同媒体的信息隐藏技术。文献中大多数无损信息隐藏都是针对图像载体设计的，也有少量针对视频和音频的。

按照嵌入域分类，无损信息隐藏方法主要可分为原始域（包括空域、时域和时空域）、变换域和压缩域方法。原始域算法和变换域算法所指的原始载体都是未压缩的原始载体，只不过变换域采用的是可逆整数变换。而压缩域算法所指的原始载体是压缩格式的图像或视频，如 JPEG 图像或 MPEG 视频。

按照鲁棒性分类，无损信息隐藏方法分为脆弱无损隐藏方法和鲁棒无损信息隐藏方法。大多数学者关注的无损信息隐藏技术是脆弱性的技术，伪装对象的轻微改变就会影响隐藏信息的提取以及原始载体的恢复。脆弱无损信息隐藏主要用于无损内容认证，即用来验证多媒体的真实性和完整性，若完成篡改检测之后发现无任何篡改（甚至对一些篡改可以自动修复），则可以完全恢复原始载体。鲁棒无损信息隐藏主要用于版权保护和交易跟踪，它需要同时实现无损和有损环境下水印的嵌入和提取，即无损环境下不失真地恢复载体对象和秘密信息，以及在受到攻击时尽可能恢复水印。

5.3 无损信息隐藏系统的框架和性能评价

5.3.1 无损信息隐藏的框架模型

无损信息隐藏方法模型框图如图 5.1 所示，主要由载体对象、秘密信息、嵌入算法、伪装对象、提取 / 恢复算法以及加密技术等主要部分组成。

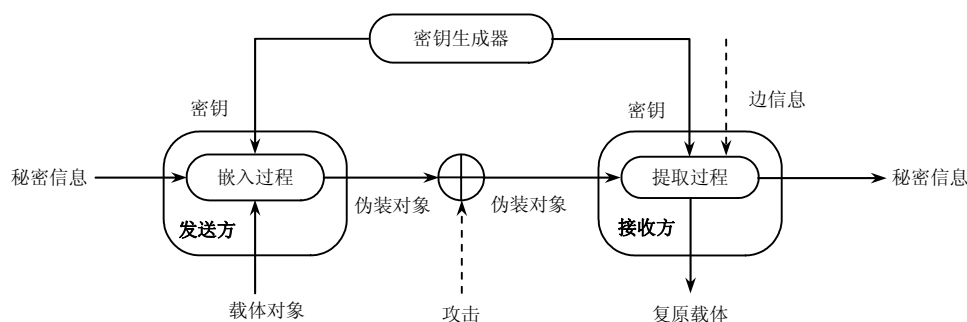


图 5.1 无损信息隐藏框架模型

1. 载体对象 C

用于隐藏信息的载体，可以是文本、图像、音频或视频信息。一般考虑人类的听觉、视觉感知系统，从载体对象中提取出对人耳、人眼不敏感的部位，作为嵌入位置，从而使伪装对象具有透明性而秘密信息不易被感知和检测。

2. 秘密信息 M

指需要被隐藏在其他载体中的信息，可以是消息、文本、图像、音频或视频等。嵌入的秘密信息将在提取过程中被提取出来，但是由于伪装对象在公共信道中传输时有可能受到隐藏分析的检测或攻击，嵌入的秘密信息有可能不能被完全正确提取。

3. 嵌入算法 Em

主要负责将秘密信息嵌入到载体对象的可修改位置中，嵌入算法应具有较高的嵌入效率或具有较大的嵌入容量，并具有较好的安全性等。

4. 伪装对象 S

嵌入过程的输出，指将秘密信息隐藏在载体对象后的结果。伪装对象应与载体对象具有相同的形式、具有较好的不可感知性。

5. 提取/恢复算法 Ex

主要负责从伪装对象中提取出所隐藏的秘密信息，并且能精确地恢复原始载体。

6. 加密算法 En

主要保证所嵌入信息的安全性，可以采用多级加密，秘密信息在嵌入前可先进行初始加密，如用 AES，RSA 等；嵌入算法在嵌入时可同时提供另一个密钥作为二级加密。

5.3.2 无损信息隐藏算法的评价

目前衡量无损信息隐藏方法的评价标准主要考虑该方法嵌入信息的容量和嵌入信息后的感知质量。但是嵌入信息的容量和嵌入信息后的视觉质量始终是一对矛盾，增加嵌

入容量常会降低感知质量，反之亦然。通常只能根据需求的不同有所侧重，在嵌入容量和感知质量寻求一个平衡，使一方得以较好的满足，而另一方做些让步。下面介绍一些主要的评价指标。

1. 不可感知性

不可感知性也叫透明性，是指人的肉眼或听觉无法感知伪装对象中的秘密信息，即在载体对象中嵌入了秘密信息后，多媒体的感知质量没有下降。通常用视觉质量或听觉质量来描述伪装对象和载体对象之间的差别，从而可以间接描述不可感知性。在实际应用中，常从两个方面来测试不可感知性，即客观评价标准和主观评价标准。在第 2 章中已经介绍了针对音频质量的 MOS 主观得分和分段信噪比的客观评价方法，以及针对图像的主观质量评分和峰值信噪比的客观评价方法，在此不再赘述，请参看 2.3.1 节。

2. 可逆性

可逆性是指含秘密信息的图像在接收端能够完全还原为原始载体，这是可逆信息隐藏技术的核心。可逆性保证了接收方从接收到的伪装对象中，得到隐藏的秘密信息之后，或根据水印完成图像的可靠性和完整性鉴定之后，能将伪装对象还原成为原始载体对象，而无任何失真。

需要注意的是，在空间域修改像素值或变换域修改系数时所遇到的最大障碍就是像素值的溢出问题。在可逆数据隐藏技术中，这个问题变得尤为严重，如果像素值发生溢出则会使得图像中出现噪声点，而且这些噪声点不能被消除，从而导致算法不可逆。若算法不考虑该问题，有可能对于某些载体对象不会出现像素值溢出问题，而对另外一些载体对象可能会出现该问题。因此，可逆性的测量可以基于大量载体对象的嵌入试验，假设一共进行了 N 次试验，在不受到任何攻击的情况下，从相应伪装对象去掉秘密信息后能够完全恢复原始载体的一共有 N' 次，则可逆性 P_{re} 可以定义为

$$P_{re} = \frac{N'}{N} \quad (5.1)$$

3. 鲁棒性

鲁棒性是指伪装对象 S 对一些常见信号处理操作（例如滤波操作，解压缩，各种几何变换等）具有的抵抗能力。对于鲁棒性好的算法，在提取端，秘密信息 M 仍然能够被检测出来且大部分恢复。测试一个系统的鲁棒性，通常模拟攻击者会根据系统采用的嵌入算法以及系统的实际应用环境来确定攻击方法，在互联网上也可以获得很多关于水印系统的检测和攻击工具，常见的有：StirMark、CentiMark 等。当然，对于脆弱无损信息隐藏系统，我们不需要鲁棒性或只是需要部分鲁棒，脆弱性和认证能力才是需要的。

信息隐藏系统的鲁棒性客观测试指标一般用误码率（Bit Error Rate, BER）来描述在提取数据时信道噪声的干扰对信息提取的影响。如果秘密信息是二值图像，还可通过采用归一化汉明距来判断提取的秘密信息和嵌入的秘密信息的相似性，定义如下

$$\rho_{HD} = \frac{1}{L} \sum_{i=1}^L M_i \oplus M'_i \quad (5.2)$$

其中， M_i 和 M'_i 分别表示嵌入的秘密信息和提取的秘密信息， L 是秘密信息的长度。一般是 ρ_{HD} 越大，系统的鲁棒性也越强。

4. 嵌入容量

这是一个很重要的系统性能评估参数，根据用户的不同应用需求，可能对系统的嵌

入容量的要求也不一样。例如，在用于版权保护的信息隐藏系统中，一般就在图像载体中嵌入一个比特就可以实现其需求，但是对于多媒体作品的发行和追踪，由于在嵌入不同的隐秘信息时，都需要复制其秘密信息，所以需要的嵌入容量就比较大。

设秘密信息 M 的数据量为 L 比特，则

$$L = \log_2 |M| \quad (5.3)$$

其中， $|M|$ 表示秘密信息集合的阶，即秘密信息集合中不同元素的个数。把载体对象是一幅图像，则可以看成一个离散的二维矩阵，设矩阵的大小为 $P \times Q$ ，则嵌入容量可以用嵌入率来表示如下

$$R_{em} = \frac{L}{P \times Q} (\text{bits/sample}) \quad (5.4)$$

5. 安全性

一个信息隐藏系统如果要推广到商业应用中，在设计系统时安全性是绕不开的因素。由于常见的嵌入算法都是公开的，此时算法将依靠密钥的空间大小来确保信息隐藏系统的安全性。在对系统的安全性进行测试时，一般从两个方面来评估：算法的计算复杂度和算法被破解所需要的时间。一个优秀的信息隐藏算法在设计时，秘密信息、数据的编码方法、数据流的嵌入位置等安全性都需要考虑在内。

6. 检测器性能

信息隐藏技术不管应用于什么领域，秘密信息的检测和提取算法都应该拥有良好的检测性能。在秘密信息检测端常见的误差有两种：漏检误差和虚警误差。漏检误差是指秘密信息存在于伪装对象中，但是检测器不能检测到其存在；虚警误差是指秘密信息不存在伪装对象中，但是检测器却能检测到伪装对象中含有秘密信息。

另外，对于认证应用场合，检测端应该不需要原始载体参与，即盲检测。以图像认证为例，因为认证水印是用来检测接收方得到的图像真实性的，如果接收方已确知原始图像，那么图像真实性的问题就不存在了；其次，有一些应用背景下根本没有原始载体，如可信赖的数码相机，为保证照片的真实性，就需要在照片拍摄的过程中自动嵌入水印，否则无法实现真实性的鉴定。所以，认证水印的检测不需要原始载体是必须的。

7. 认证性能

具有认证功能是认证型无损信息隐藏技术的本质要求，即在伪装对象受到篡改攻击时，认证系统应该可以比较可靠地发出篡改提示，并在有需要时，能准确地实现篡改定位，且可靠地证实伪装对象其他部分的真实性和完整性，甚至可以恢复篡改位置的原始载体信息。一个完备的认证水印系统应该满足如下要求：① 能准确地判断数字媒体内容是否被篡改；② 如果数字媒体内容被篡改，能够有效描述被篡改程度；③ 在无法得知原始媒体内容或其他与真实信号内容相关的信息的条件下，认证系统最好能够判断可能的篡改操作的具体类别；④ 可以实现篡改定位，并能进行有效的篡改部分的恢复。

5.4 空域无损信息隐藏技术

由于大多数无损信息隐藏技术都是针对图像载体的，所以后续的内容只介绍一些典型的图像无损信息隐藏方法。最早关于可逆信息隐藏的思想是 1997 年 Barton 提交的专利。Barton 通过无损压缩原始数据流挤出一定的空间以嵌入有用信息。1999 年 Honsinger

等人在美国柯达公司发表了一个专利，这项专利利用无损的嵌入信息方法，在图像空域中采用模加的方法来嵌入隐藏信息，但是嵌入的数据量较少，而且会由于隐藏信息的嵌入，产生数据的溢出，从而造成椒盐噪声，影响图像的视觉质量。2001 年，Fridrich 又提出一种针对图像的无损信息隐藏方法^[73]，利用分组压缩替换的方法来实现无损的嵌入信息。此后，越来越多的空域无损信息隐藏方法提出。下面介绍现有的几种典型方法。

5.4.1 基于无损压缩替换的无损信息隐藏

第一类经典的无损信息隐藏算法是利用无损压缩原理来对载体媒体的冗余进行压缩，将空出的比特空间用于嵌入秘密信息。Fridrich 等提出的**无损位平面压缩**（Lossless Bit-plane Compression）^[73]和 R-S 算法^[74]以及 Celik 等提出的**广义最低有效位无损压缩**（General LSB Lossless Compression）^[75]算法就属于这一类。下面简要介绍这三种算法，实际上，最后一种基于广义 LSB 无损压缩的核心思想同 Fridrich 的无损位平面压缩方法类似，只是将被无损压缩的对象由原有的最低 n 个位平面扩展为最低的 L 个幅值电平。

1. 基于无损位平面压缩的无损信息隐藏

Fridrich 在文献^[73]中提出一种利用 JBIG 无损压缩比特平面的方法实现无损信息隐藏，这个思想简单有效。该方法先查看载体图像的第 5 个位平面（定义最低位平面为 LSB 平面），利用 JBIG（Joint Bi-level Image experts Group）压缩得到冗余 R ，单位为比特，即

$$R = \text{像素总数目} - \text{压缩后的数据长度} \quad (5.5)$$

假设要嵌入的秘密信息长度为 L ，那么先查看载体图像的第 5 个位平面得到的冗余 R 是否大于或等于 L 。如果 R 大于或等于 L ，那么接下去查看更低的一个位平面，即第 4 个位平面；否则接下去查看更高的一个位平面，即第 6 个位平面。如此反复操作直到不能再向更低的位平面继续为止。显然，为了使水印嵌入失真最小，应该选择最后确定能提供足够冗余的最低的那个位平面为秘密信息的嵌入平面，称作**关键位平面**（Key Bitplane）。对此平面用 JBIG 进行无损压缩，在空出的比特空间内嵌入秘密信息，嵌入信息后新计算的冗余 R' 一定比原来的冗余 R 要小，即 $R' < R$ ，但无法保证 $R' < L$ ，不过可以保证比此关键位平面更低的下一个比特平面肯定满足 $R' < L$ 。在进行认证和图像恢复时，首先查找到满足 $R' < L$ 的最高的位平面 i ，由上面的信息嵌入过程可以看到，关键位平面只可能是位平面 i 或者 $i+1$ ，那么我们对这两个位平面都进行水印提取，只要其中一个平面能正确提取水印信息并进行认证，就可以宣称载体图像可被成功认证，否则可宣称认证失败。这种利用压缩位平面进行无损信息嵌入的方法思想简单、实现容易，但数据容量较小，通常只能作认证之用。

2. R-S 算法

为了增加数据容量，Fridrich 在文献^[74]中提出了一种类似拼贴算法像素分组思想的无损信息隐藏算法。该算法首先将原始图像像素分组，例如将邻近的 n 个像素 c_1, c_2, \dots, c_n 分为一组，标记为 C ，然后利用一个区分度函数计算每个分组的区分度 $f(c_1, c_2, \dots, c_n)$

$$f(c_1, c_2, \dots, c_n) = \sum_{i=1}^{n-1} |c_{i+1} - c_i| \quad (5.6)$$

实际上这个区分度衡量的是每个像素分组的“平滑”程度，称作“正常性”

(Regularity)。接着引入一个幅度为 A 的翻转操作 F ，即将灰度值进行可逆的翻转，翻转后的值同原值相差 A ，并满足 $F(F(c))=c$ ， $c \in P$ ， $P=\{0, 1, \dots, 255\}$ 。翻转幅度 A 的定义如下

$$A = \frac{1}{|P|} \sum_{c \in P} |c - F(c)| \quad (5.7)$$

例如， $A=1$ 时，可定义翻转操作为 $0 \leftrightarrow 1, 2 \leftrightarrow 3, \dots, 254 \leftrightarrow 255$ ； $A=2$ 时，可定义翻转操作为 $0 \leftrightarrow 2, 1 \leftrightarrow 3, 4 \leftrightarrow 6, 5 \leftrightarrow 7, \dots$ 。显然，这种翻转操作的目的在于以一种可逆的方式打乱原有像素值的相关性，可看作一种可逆地添加微小噪声的操作。对于普通自然图像而言，引入这种翻转操作会使绝大多数像素分组的区分度增加，定义这些像素分组为正常分组 G_R ，即 $C \in G_R \Leftrightarrow f(F(C)) > f(C)$ ；若区分度减小，定义为奇异分组 G_S ，即 $C \in G_S \Leftrightarrow f(F(C)) < f(C)$ ；若区分度不变，定义为不可用分组 G_U ：即 $C \in G_U \Leftrightarrow f(F(C)) = f(C)$ 。

在水印嵌入过程中，可将水印比特“1”对应“使像素分组 C 的状态为 G_R ”，“0”对应“使像素分组 C 的状态为 G_S ”，而对所有状态为 G_U 的像素分组 C 不嵌入任何数据，也不改变其状态。由于这种翻转操作是可逆的，因而只要知道原始载体图像的像素分组状态，就能完全恢复原来图像分组的所有像素值。在此算法中，原始图像像素分组状态属于边信息，可经无损压缩记录为一个 R-S 矢量 V ，并可将其同秘密信息一起嵌入到载体图像中。因此，此算法的嵌入容量为

$$\text{Capacity} = N_R + N_S - |V| \quad (5.8)$$

其中 N_R 和 N_S 分别为原始载体图像正常分组和奇异分组的数目，即可嵌入数据的像素分组数目， $|V|$ 为压缩后的原始载体图像像素分组状态所占的长度。可以看出，此算法的嵌入容量主要取决于 $|V|$ 的大小，只有当原始载体图像的正常分组数目和奇异分组数目相差很大时， V 才有可能被压缩得很小，因而从原理上说，R-S 算法属于无损压缩替换方法的范畴。

实验结果表明， $n=4$ 的像素分组的嵌入容量最高，因为当 n 过小时， G_R 分组和 G_S 分组的数目相差太小， $|V|$ 会很大；而当 n 过大时， G_R 分组和 G_S 分组的总数目太小， $N_R + N_S$ 也就受到限制。

3. G-LSB 算法

Celik 等的 G-LSB 方法^[75]首先对载体图像 C 中的各像素 c 进行量化

$$Q_\Delta(c) = \Delta \cdot \left\lfloor \frac{c}{\Delta} \right\rfloor \quad (5.9)$$

其中 c 为各像素的灰度值， Δ 为所选的量化参数。然后用载体图像中的各像素的灰度值 c 减去这个量化值，得到剩余量 x

$$x = c - Q_\Delta(c) \quad (5.10)$$

再将所有 x 的数值用基于上下文自适应无损压缩方法 (Context-based Adaptive Lossless Image Codec, CALIC) 进行压缩，然后把压缩后的数据与需要嵌入的秘密信息 M 合并成为最终的待嵌入信息 W ，由下式完成信息嵌入

$$S = Q_\Delta(C) + W \quad (5.11)$$

其中， C 为原始载体，而 S 为嵌入隐藏信息后的伪装对象。而提取信息时则采用下式提取隐藏的信息

$$W' = S' - Q_\Delta(S') \quad (5.12)$$

其中 S' 是可疑的伪装对象， W' 是提取的信息。若 S 未受攻击，即 $S'=S$ ，则 $W'=W$ ，这样 W' 中可以剥离出秘密信息 M ，而将剩余信息经过解压缩过程获得各像素点的 x 值，

最后利用 x 和 Q_{Δ} 的逆过程来恢复原始载体的各像素点 c 。

5.4.2 基于模加的无损信息隐藏

基于模加的鲁棒无损信息隐藏算法最早出现于 1999 年柯达公司的 Honsinger 等人申请的专利, 之后 Fridrich 等人将这个思想引申到了 DCT 变换域。模加思想可以实现可逆, 是因为非自适应的加性水印信号可以被无损失地从像素值上直接减去, 只要知道可能嵌入的水印模式且不会出现像素值溢出的问题, 这个模加的操作就是完全可逆的。这种模加的方式用来嵌入的都是载体媒体的认证信息。由于算法鲁棒, 这里采用可逆水印术语来描述无损信息隐藏, 其水印嵌入和认证过程分别描述如下。

1. 鲁棒可逆水印的嵌入

鲁棒可逆水印嵌入过程如图 5.2 所示, 具体描述如下。

步骤 1: 令 C 为原始待认证图像, 计算其哈希 $H(C)$;

步骤 2: 选择一种加性、非自适应性的鲁棒水印技术, 通过一个密钥 K 产生一个水印模式 W , 使得 W 的有效载荷是 $H(C)$, 这里要求水印模式 W 只是密钥 K 和有效载荷 $H(C)$ 的函数, 即 $W=f(K, H(C))$;

步骤 3: 使用模加操作 \oplus 将水印模式 W 加到 C 上, 产生待认证图像 $S=C\oplus\alpha W$, 其中 α 为水印强度控制系数。

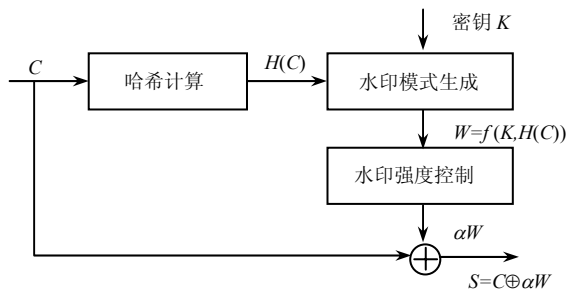


图 5.2 鲁棒可逆水印的嵌入过程

2. 基于鲁棒可逆水印的认证

基于鲁棒可逆水印的认证过程如图 5.3 所示, 描述如下。

步骤 1: 从可疑伪装对象 S' 中提取水印比特串 H' (有效载荷);

步骤 2: 通过密钥 K 和提取的比特串 H' 生成水印模式 $W'=f(K, H')$;

步骤 3: 将 W' 从 S' 中减去, 得到 $C'=S'-\alpha W'$;

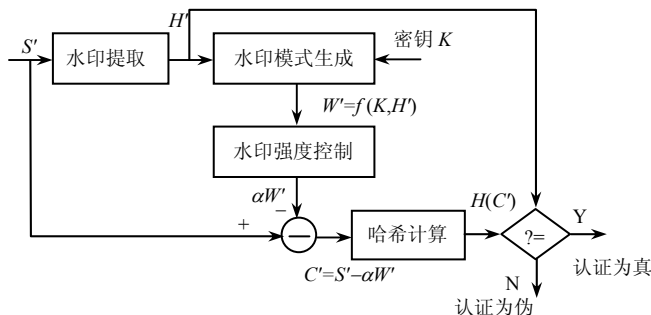


图 5.3 基于鲁棒可逆水印的认证过程

步骤 4: 计算哈希 $H(C')$, 同先前提取的哈希 H' 相比较, 如果相符合, 图像可被认证为真, 即 C' 为原始未遭改动和攻击的图像; 否则认证失败, 载体图像已遭到改动或攻击。

在上面的水印可逆嵌入和认证的过程中, 模加操作定义如下

$$i \oplus k = L \cdot \left\lfloor \frac{i}{L} \right\rfloor + (i + k) \bmod L \quad (5.13)$$

其中, i 和 k 为参与模加的两个整数, L 为模长, \bmod 为取余数部分的操作。通过这种模加操作, 若 $i+k$ 恰好等于模长 L 的整数倍, 则 $(i+k) \bmod L=0$, 那么模加的结果 $i \oplus k$ 同 $i+k$ 差值的绝对值 D 正好相差一个模长 L 。例如, 在模长 $L=16$ 的情况下, 设 i 的动态范围为 $[0,255]$, $k=1$, 则此模加操作的结果为 $0 \rightarrow 1, 1 \rightarrow 2, \dots, 15 \rightarrow 0, 16 \rightarrow 17, 17 \rightarrow 18, \dots, 31 \rightarrow 16, \dots$ 。显然, 如果模长 L 越小, 这种模加操作在边界 ($i+k$ 恰好等于模长 L 的整数倍的情况) 引起的失真越小; 但相应地, 模长 L 越小, 在整个动态范围内 (如 8 位灰度图像的动态范围 $[0,255]$) 出现模加失真的像素点也越多。对于一幅给定图像, 尽可能减小模加带来的像素值失真, 需要根据具体情况对 L 进行选择: 若 k 值小, 且图像边缘像素值 (靠近灰度 0 和 255 的像素值) 较少, 则可采用大 L 值, 如 256; 否则, 应考虑采用小 L 值, 以免引起椒盐噪声之类的严重视觉失真。另外, 需要注意, 在认证步骤①中, 提取水印使用的是噪声序列作相关运算的方法, 因而能使水印提取过程具有鲁棒性。

5.4.3 基于差值扩展的无损信息隐藏

差值扩展 (Difference Expansion, DE) 方法用于无损信息隐藏最早由 Tian 提出^[76]。差值扩展的基础是**比特移位** (Bit-shifting) 技术, 具有较大的数据容量。下面首先介绍比特移位, 然后介绍 Tian 的 DE 方法。

1. 比特移位技术

设有一长为 N 的时间离散信号 $C=\{c_i\}$, $c_i \in \{0, 1, \dots, 2^m-1\}$, $i \in \{1, 2, \dots, N\}$, 其中 m 为表达一个采样所需的比特数。注意到, 若信号样值 c_i 的最大值 c_{\max} 不大于最大允许值, 即

$$c_{\max} \leq 2^{m-1} - 1 \quad (5.14)$$

那么这个整数值 c_i 的幅度可被放大至原来的 2 倍而不会引起数值溢出失真, 这个幅值放大 2 倍的操作相当于将所有的比特位左移 1 位, 即

$$s_i = 2c_i \quad (5.15)$$

例如 $c_i=3$ (011B), $s_i=6$ (110B)。可以看到, 经过比特移位后的整数其 LSB 位为 0, 结果总是偶数。此时的 LSB 位便可用来嵌入“0”“1”的二值水印数据。显然如果信号样值最大值 c_{\max} 满足

$$c_{\max} \leq 2^{m-p} - 1, \quad 1 \leq p \leq m-1 \quad (5.16)$$

则 x_i 值可以左移 p 位而没有任何数值溢出失真。

若要将比特移位技术应用到无损信息隐藏算法中, 有两个问题需要解决: 第一, 如果信号样值最大值 c_{\max} 大于最大允许值, 即 $c_{\max} > 2^{m-p} - 1$, 那么比特移位的结果必然造成数值溢出。由于表示每个样值的二进制比特位数是有限的, 数值溢出必然产生不可逆的失真, 因此利用比特移位技术实现无损信息隐藏必须以式 (5.16) 为前提; 第二, 对于满足式 (5.16) 的 c_i , 1 次比特左移位所产生的误差恰好为 c_i 的幅值绝对值, 即

$$d = |s_i - c_i| = |2c_i - c_i| = |c_i| \quad (5.17)$$

由于每次比特移位所能提供的水印嵌入空间都是 1 比特（即 s_i 的 LSB 位），我们自然希望比特移位的对象 c_i 的幅值绝对值越小越好，因此利用比特移位技术实现无损信息隐藏要选择幅值绝对值小的 c_i 为比特移位对象或进行预处理以减小 x_i 的幅值绝对值为原则。

2. 基于差值扩展的无损信息隐藏

由上面的描述可知，对于 1 次比特左移而言，获得的嵌入空间总为 1 比特，因此希望原始幅值 c_i 能够尽量小，使得由于比特移位操作造成的误差也能相应地减小。如果直接在图像的空间域上作比特移位，幅值乘 2 造成的误差在视觉上无法让人接受，而且会出现大量数值溢出的情况，此时若在边界（如 256 灰度的 0、255 灰度）上硬截断，则会造成大量像素值集中于上下边界，而且这种硬截断是不可逆的。为此，Tian 提出了差值扩展技术，实际上是对整数 Haar 小波（或称 S-变换）的高频系数进行比特移位来进行数据嵌入。因为对于邻近像素间存在大量冗余的自然图像而言，其整数 Haar 小波变换的高频系数的幅值动态范围同像素值的动态范围相比大大减小，而且其幅值分布围绕着 0 值呈能量集中式的广义高斯分布，减小的幅值绝对值非常有利于进行基于比特移位的可逆数据嵌入。

Tian 所用到的灰度图像的 Haar 整数小波变换定义如下。对于邻近灰度像素对 $\{x, y\}$, $0 \leq x, y \leq 255$ ，定义正变换为

$$\begin{cases} l = \left\lfloor \frac{x+y}{2} \right\rfloor \\ h = x - y \end{cases} \quad (5.18)$$

其中， l 被称为整数均值， h 被称为整数差值。其对应的逆变换为

$$\begin{cases} x = l + \left\lfloor \frac{h+1}{2} \right\rfloor \\ y = l - \left\lfloor \frac{h}{2} \right\rfloor \end{cases} \quad (5.19)$$

可以证明，这个整数 Haar 小波变换是一个双射（Bijection），能满足无损信息隐藏的要求。由于自然图像相邻像素间冗余很大，从统计上来说，通常经过此整数变换后其高频系数（即整数差值 h ）的幅值绝对值要比一般的像素值小得多，因而在进行比特移位嵌入数据的同时能减小幅值改变带来的图像质量下降。如果对 h 进行比特移位，则有

$$h' = 2h + b \quad (5.20)$$

其中 b 为嵌在 LSB 位上的“0”和“1”的水印数据。为了不引起像素溢出，反变换得到的 x 、 y 值应该被限制在 $[0, 255]$ 范围之内，否则水印提取和图像恢复就不再可逆。因此，还应有以下限制条件

$$0 \leq l + \left\lfloor \frac{h+1}{2} \right\rfloor \leq 255 \text{ 且 } 0 \leq l - \left\lfloor \frac{h}{2} \right\rfloor \leq 255 \quad (5.21)$$

因为 l 和 h 都是整数，上式等价于

$$|h| \leq 2 \cdot (255 - l) \text{ 且 } |h| \leq 2l + 1 \quad (5.22)$$

或者等价于

$$\begin{cases} |h| \leq 2 \cdot (255 - l) & 128 \leq l \leq 255 \\ |h| \leq 2l + 1 & 0 \leq l \leq 127 \end{cases} \quad (5.23)$$

显然，要避免比特移位后像素值溢出现象的发生，就必须有

$$|h'| \leq \min \{2 \cdot (255 - l), 2l + 1\} \quad (5.24)$$

基于上面描述, 给出 h 值可扩展 (Expandable) 和可交换 (Changeable) 的定义如下。

定义 5.1: 称差值 h 在整数均值 l 下是可扩展的, 若对 $b=0$ 和 $b=1$ 都能满足

$$|2h + b| \leq \min \{2 \cdot (255 - l), 2l + 1\} \quad (5.25)$$

显然, 能满足式 (5.25) 的 h 值可被用作比特移位, 即差值扩展。

定义 5.2: 称差值 h 在整数均值 l 下是可交换的, 若对 $b=0$ 和 $b=1$ 都能满足

$$\left| 2 \cdot \left\lfloor \frac{h}{2} \right\rfloor + b \right| \leq \min \{2 \cdot (255 - l), 2l + 1\} \quad (5.26)$$

关于可扩展和可交换, 有如下关系。

- (1) 对于一个可交换差值 h , 若改变它的 LSB 位, 改变后的差值仍然可交换;
- (2) 如果一个差值 h 是非可交换的, 那么 x 必是 0 或 255, 且 y 是一个奇数;
- (3) 如果 x 是 0 或 255, y 是一个奇数, 那么差值 h 是非可交换的;
- (4) 可扩展的差值总是可交换的;
- (5) 对于可扩展的 h , 扩展后的差值 h' 是可交换的。
- (6) 如果 $h=0$ 或 -1 , 可扩展和可交换的条件是等价的。

有关证明可见文献[76], 这里不再给出。

根据 h 可交换性与可扩展性的差异, 可将所有的差值 h 在相应的整数均值 l 下分为四个集合 EZ、EN、CN 与 NC。① EZ: 包括所有可扩展的 $h=0$ 和 $h=-1$; ② EN: 包括所有可扩展的 $h \notin \text{EZ}$; ③ CN: 包括所有可交换的 $h \notin (\text{EZ} \cup \text{EN})$; ④ NC: 包括所有非可交换的 h 。每一个差值 h 都会归属于以上这四个集合中唯一的一个。注意, 因为所有可扩展的 h 同时也是可交换的, 所以整个可交换的差值集合为 $\text{EZ} \cup \text{EN} \cup \text{CN}$ 。考虑到 EN 集合中的 h 经过比特移位的扩展后, 有可能变为可交换但非可扩展 (即归入 CN), 那么在 EN 中的 h 上嵌入水印后, 得到的可交换但非可扩展的 h' 就无法同原始 CN 集合中的 h 相区别, 造成含水印图像无法被解码, 因而无法提取水印与恢复图像。为解决这一问题, Tian 使用了一种称作位置地图 (Location Map) 的附加信息记录已被选择进行差值扩展的那些差值 h 。在算法中, 所有 EZ 中的 h 均被选择进行差值扩展, 而 EN 中的 h 则根据数据容量的需要被部分或全部选择进行差值扩展, 其中所有被选择的 h 定义为集合 EN1, 未被选择的定义为 EN2。在位置地图中, 用“0”、“1”比特来记录每一个原始 h 的状态: 若原始的 h 在 EZ 和 EN1, 则对位置地图中对应 h 所产生的像素对位置分配“1”; 若原始的 h 在 EN2, CN 或 NC 中, 则分配“0”。这样, 使用位置地图, 解码时由原始属于 EN 的 h 扩展而来的现属于 CN 的 h' 就会和原始属于 CN 的 h 区别开来, 继而能进行准确的水印提取和图像恢复。

在该算法中, 位置地图比特串 L 将连同水印比特串 W 被存储在所有可交换的差值 h ($\text{EZ} \cup \text{EN} \cup \text{CN}$) 的 LSB 位上, 因为可交换的 h 的 LSB 位的改变不会造成像素溢出。但需要注意, 对于已扩展的 h ($\text{EZ} \cup \text{EN1}$) 而言, 其原始 LSB 位在比特移位中能被完整地保留 (被左移了一位), 而其余可交换的 h ($\text{EN2} \cup \text{CN}$) 在 LSB 位的交换过程中原始的 LSB 信息则会被丢弃, 因此在向所有可交换的 h ($\text{EZ} \cup \text{EN} \cup \text{CN}$) 的 LSB 位存储信息之前, 必须提前记录可交换的 h ($\text{EN2} \cup \text{CN}$) 的原始 LSB 信息 O , 否则原始图像在解码阶段就不可能完全恢复了。综上可以看到, 在此算法中, 需要进行可逆嵌入的数据 B 由三部分组成: 位置地图信息 L (可由 JBIG 算法进一步压缩), 原始 ($\text{EN2} \cup \text{CN}$) 中 h 的 LSB 位信息 O , 以及水印信息 W 。表 5.1 给出了差值 h 的分类和数据嵌入状态。

表 5.1 差值分类与数据嵌入状态

分 类	原始集合	原始差值	位置地图	新差值
可交换	EZ 或 EN1	h	1	$2h+b$
	EN2 或 CN	h	0	$2\left\lfloor \frac{h}{2} \right\rfloor + b$
非可交换	NC	h	0	h

在水印提取和图像恢复（即解码）阶段，首先从所有可交换差值的 LSB 位提取所嵌入的比特串 B ，并从 B 中解码得到位置地图 L ，原始 LSB 位信息 O ，以及水印信息 W 。将水印信息 W 完整地提取后，继续进行图像的恢复。根据位置地图 L ，能得到经扩展的差值 h' ($EZ \cup EN1$) 的位置，对其进行右比特移位，原始差值 h 便能恢复；对于其他可交换的差值 ($EN2 \cup CN$)，将其 LSB 位恢复为 O 所记录的比特串。这样，原始图像就被完全恢复了。

对自然图像而言，大约 99% 以上的像素对是可扩展的，因而利用上述差值扩展技术，无损信息隐藏算法可获得非常大的数据容量。除此之外，Thodi 和 Rodriguez 在文献[77]中提出一种称为预测误差扩展的可逆数据隐藏算法，基本思想是首先预测某个像素 c 的修改后大小 c' ，那么预测误差的值就为 $P_e = c - c'$ ，而二值水印信息就是通过预测误差 P_e 的扩大而嵌入的。预测误差扩展的方法要优越于 Tian 的算法，由于前者得到的可供水印嵌入的空间较为充沛，理论上可以达到 1 比特/像素。限于篇幅，在此不再赘述。

5.4.4 基于直方图移位的无损信息隐藏

Ni 等人于 2003 年提出了一种基于直方图移位的可逆水印新技术^[78]，这种技术将图像的空间域直方图的部分区域进行整体平移，人为制造冗余以调制需要嵌入的水印数据。我们用图 5.4 来说明这个方法，其中的三幅直方图分别表示移位前（原始图像直方图）、移位后、以及嵌入水印后（含水印图像直方图）的状态。在所有三幅直方图中，横坐标 0, 1, 2, ..., 255 为灰度值， N 为原始图像空间域直方图（即统计所有像素而成的直方图）中的峰值（即图像中出现次数最多的像素数目），其对应的灰度值为 P ，称作**峰点**（Peak Point），相应的 P 之右第一个幅值为 0 的灰度值为 Z ，称作**零点**（Zero Point）。在进行直方图移位之前，对当前直方图找到峰点 P 和零点 Z ，然后进行移位，即将灰度值区域 $[P, Z-1]$ 内的直方图部分向右平移一个灰度刻度，也就是对原始图像所有在 $[P, Z-1]$ 区域内的像素值加 1，而使原来的峰点 P 被空出，而原来的 $P+1$ 灰度刻度上的幅值则变成了直方图峰值 N 。接下来，可以将幅值 N 根据水印数据“0”或“1”调制到 P 或 $P+1$ 上，也就是将直方图移位后的图像中值为 $P+1$ 的像素重新赋值：若需要嵌入的水印位为“0”，则将此像素的值 $P+1$ 减 1 得到 P （即恢复到原始图像的像素值状态）；若要嵌入的水印位为“1”，则保持移位后的新像素值 $P+1$ 不变。这样便完成了整个水印嵌入的过程。注意到，只要知道峰点 P 和零点 Z ，上述基于直方图移位的水印嵌入过程就是完全可逆的。当想要提取水印恢复原始图像时，首先获得峰点 P 和零点 Z 的具体位置，然后由嵌入过程可知，此时所有像素值为 P 的像素所调制的水印数据位为“0”，而所有像素值为 $P+1$ 的像素所调制的水印数据位为“1”。提取水印信息后，只要将灰度值区域 $[P+1, Z]$ 内的所有像素的值减 1 即可完全恢复原始图像。

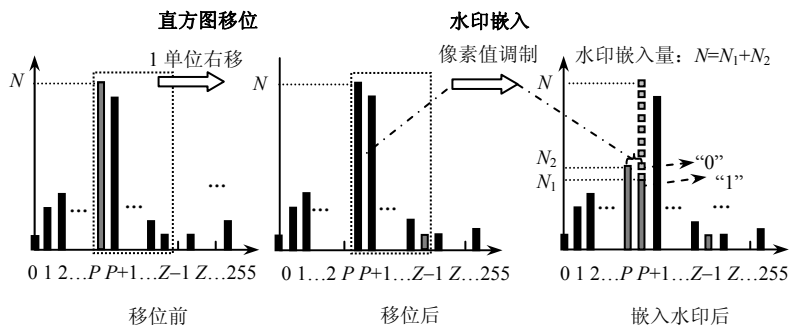


图 5.4 基于直方图移位技术的可逆水印嵌入

相对于前面提到的可逆水印算法，这种基于直方图移位技术的可逆水印算法能够在高图像质量要求下提供更大的数据容量。但因为此方法建立在空间域直方图上，数据容量容易受图像不同统计特征的影响，性能稳定性较差。图 5.5 给出了 Lena 和 Airplane 两幅灰度图像的直方图，可以看出 Airplane 的直方图的峰值要远远高于 Lena 图，显然更有利于使用这种空间域直方图移位的技术。

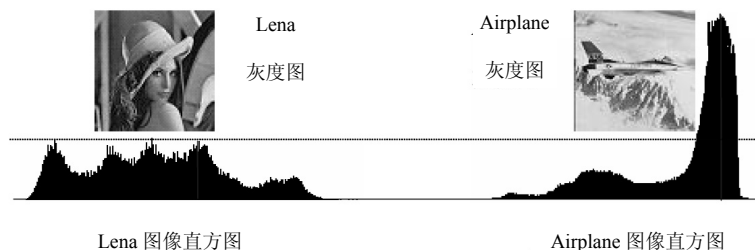


图 5.5 Lena 和 Airplane 直方图对比

另外需要说明的是，并非所有图像的直方图中都存在零点。对于不存在零点的直方图，取幅值最小的灰度刻度为零点 Z 。当然，此时需要记录原有那些处于 Z 的像素值位置信息，连同 P 和 Z 的数值作为附加信息嵌入在图像的某些地方，在提取水印和恢复原始图像之前，这些附加信息应被正确提取。显然这些附加信息嵌入的位置不能再被水印数据所覆盖，通常可以用一个密钥来控制附加信息嵌入的位置，而将直方图移位和水印嵌入操作在除这些附加信息嵌入位置之外的其他像素上进行。

5.5 变换域无损信息隐藏技术

从概念上讲，5.4.3 小节介绍的整数 Haar 小波变换域高频系数（即差值）扩展技术也属于整数变换域的数值扩展技术，但是由于差值扩展技术应用对象仅仅局限在相邻像素对上，其效果更接近于空间域预测解相关的处理方式，因而将其归为空间域数值扩展技术。本节将介绍基于 8×8 整数余弦变换（Integer DCT）域和整数小波变换（Integer DWT）域的无损信息隐藏技术，主要技术手段不外乎数值扩展和直方图移位等，下面分别加以介绍。

5.5.1 整数 DCT 变换域数值扩展技术

根据前面的描述可知，比特移位会造成数值幅值 2 倍的扩大，那么在获取相同的数据容量的情况下，只有被扩展的数值幅值越小，水印嵌入所造成的图像质量失真才会越

小。前面 5.4.3 节用来达到这个目的的手段是对存在高度相关性的邻近像素对作整数 Haar 小波变换, 得到的高频系数 (即整数差值) 的幅值范围在统计概念上被大大缩减, 从而适合比特移位操作。从这个角度考虑, DCT 变换具有较高的解相关效率, 应该非常适合从冗余度较大的自然图像的 DCT 变换结果获得大量小幅值的系数。实验表明, 自然图像的 8×8 DCT 变换的每个系数位置的数值分布可归纳为广义高斯分布, 而且越趋向高频位置越呈现出集中分布在 0 值附近的趋势, 并且图像冗余度越高, 系数分布形状越“窄” (即小幅值系数越多), 越有利于发挥比特移位的优点。这就给我们一个启示: 对 DCT 的高频系数使用比特移位技术应该能够获得最佳的水印嵌入性能。但是, 注意到浮点数 DCT 存在舍入误差和运算过程的累积误差, 难以满足可逆性要求, 因此在算法实现时只能考虑满足可逆性要求的整数变换。为此杨等人^[79]在德国 Duisburg 大学教授 Planka 和 Tasche 提出的一维 8 点整数 DCT^[80]的基础上, 推得了相应的 8×8 整数 DCT 变换, 用于获得整数 DCT 域的整数 AC 系数。这个 8×8 整数 DCT 是一个由 2×2 的提升矩阵的整数逼近运算的线性组合形式构成, 是一个输入输出都在整数域上的双射。实验表明, 对于一般的自然图像, 这个 8×8 整数 DCT 输出的整数结果同 8×8 浮点数 DCT 输出结果的误差主要在 $0 \sim 6$ 的范围内, 超过 6 的误差极为少见。

图 5.6 给出了基于 8×8 整数 DCT 的可逆水印算法的基本框架。其中 C 为原始图像中的某个 8×8 像素块, D 为 C 经过 8×8 整数 DCT (图中的 Int.DCT) 得到的系数块; S 为含水印的像素块, 由含水印的系数块 D_W 经过 8×8 整数逆 DCT 变换得到; D_{BS} 为 D 经过系数选择和比特左移位得到的系数块, 被选择并移位后的系数的后 p 个 LSB 位嵌入水印 W 后得到含水印系数块 D_W 。整个嵌入和提取过程中, 系数的选择都由密钥 K 来控制, 整数 DCT 变换及其逆变换都是无损的, 比特移位过程也是无损的, 从而保证了整个水印处理过程是完全可逆的。

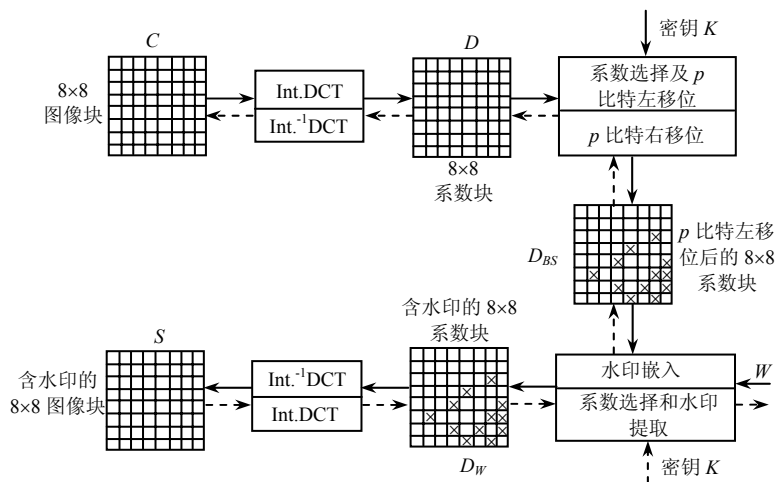


图 5.6 8×8 整数 DCT 系数块的可逆水印算法框图 (实线: 嵌入过程; 虚线: 提取过程)

然而, 在实际图像处理过程中, 图 5.6 的可逆水印处理过程完全可逆还须满足: 避免像素值溢出现象的发生。显然, 如果经过整数逆 DCT 变换得到的含水印图像块 S 中有像素超出了原先的动态范围, 如 $[0, 255]$, 这个图像块最后存储和显示的值一定不会和 S 相同, 这样在水印提取的过程中, 可逆性就丧失了。为了在实际应用过程中避免像素值溢出现象的发生, 需要先来分析像素值溢出现象可能的来源。显然, 对整数 DCT 域系数

的改动是像素值溢出的原因,而改动有两方面:第一是比特移位;第二是水印嵌入。如果能在水印嵌入之前估计出这些改动所造成的空间域的像素误差,就能提前预测哪些像素块可能会出现像素值溢出现象,从而避免对其进行水印嵌入的操作。一旦能估计出像素块中每个像素点的误差值上下限,就能在像素值的动态范围 $[0, 255]$ 内设置相应的阈值数组 TH_l 和 TH_h (均含有 8×8 个阈值,对应 8×8 的像素块),只有在动态范围 $[TH_l, TH_h]$ 内的像素块才完全没有像素溢出的可能,能被用来进行可逆水印处理,即要求原始像素块 C 满足

$$TH_l \leq C \leq 255 - TH_h \quad (5.27)$$

上面提到可以通过对像素块进行误差估计来决定对哪些像素块进行可逆水印的嵌入,而另一些块则不作处理。这会产生另一个问题:如何在水印提取的阶段准确地甄别出哪些像素块经过了修改而另一些并未遭到改动呢?因为完全存在这样的可能:原有符合动态范围 $[TH_l, TH_h]$ 条件的像素块嵌入水印后变得不再符合动态范围 $[TH_l, TH_h]$ 条件,这样就无法仅仅根据动态范围阈值条件来进行区分。一个可能的办法是利用附加信息来记录不同性质像素块的位置,类似于差值扩展方法中使用的位置地图。但在何处存储这些像素块位置信息又是一个麻烦的问题。最理想的方法是不需要任何其他先验信息,就能从任一幅含水印灰度图像中直接提取这些附加信息。这就要求这些附件信息只能被嵌入在一些具有嵌入不变性的像素块中,也就是说,要求这些像素块能够在嵌入一次水印信息后还能被准确地识别出来。显然,如果一个像素块能够经过两次比特移位和水印嵌入仍然处于式(5.27)所确定的范围内,就能直接通过虚拟的两次比特移位和水印嵌入来直接识别,并可将附加信息直接嵌入到满足此条件的像素块中。经过二次嵌入而不引起像素值溢出的像素块需要满足的动态范围阈值条件为

$$TH'_l \leq C \leq 255 - TH'_h \quad (5.28)$$

根据式(5.27)和式(5.28)可将原始像素块分为三类:分类 I 为不能满足式(5.27)的像素块,即对整数 DCT 系数修改较敏感的像素块,我们不对这一类像素块作任何水印处理操作;分类 II 为能满足式(5.27)但不能满足式(5.28)的像素块,这一类块用来嵌入水印信息;分类 III 为同时满足式(5.27)和式(5.28)的像素块,这一类块可在水印提取之前无需其他任何信息而被正确识别,可作为记载像素块分类信息(尤其是区分分类 I 和分类 III 的信息)和其他附加信息的像素块。将这种二次嵌入检测像素块分类状态的方法称作“二次测试”。而在水印提取过程中,由于经过一次水印嵌入的原分类 II 中的一些像素块很可能不再满足式(5.27),故进行式(5.27)的测试时,会随同原分类 I 的像素块被分在一起,将其标记为分类 II';另一部分原属分类 II 的像素块则依然能满足式(5.27)而随同原分类 III 被分在一起,将其标记为分类 II"。为了区分分类 II"和分类 III,将它们全部进行右移位,这样便得到了原始的像素块图像,再接着对这些被恢复的原始像素块作与嵌入过程一样的甄别,便能将分类 II'和分类 III 清晰分离开。一旦将分类 III 分离出来后,便可从中提取块分类信息,进而可准确地将分类 II'和分类 I 分离。

总结以上分析过程,整数离散余弦变换域可逆水印嵌入和提取可以归纳为图 5.7(a)和(b),其中 L 为块分类信息, C_R 和 D_R 分别为恢复后的像素块和系数块。同其他可逆水印算法相比,上面提出的基于数值扩展的整数 DCT 域可逆水印可以大大提高水印的数据容量,同时在提取阶段不需要先验信息或离线存储数据的参与。

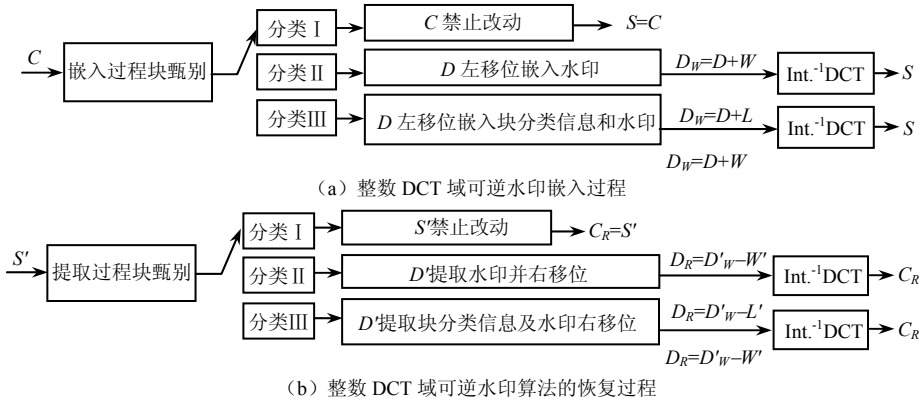


图 5.7 整数 DCT 域可逆水印算法框图

5.5.2 整数 DWT 变换域数值扩展技术

Xuan 等^[81]在杨等人提出的整数 DCT 域比特移位技术的思想上, 尝试利用整数小波 CDF(2,2)的一级分解获得了三个高频段 LH、HL、HH 的系数, 并在这些系数上进行压扩和数值扩展, 取得了较好的算法性能。注意到在这个算法中, 性能提高的关键是其压缩函数的设计。下面首先介绍其使用的 CDF(2,2)整数小波的形式, 然后给出 Xuan 等所使用的压缩函数。最后分析其压扩误差的计算, 提出一种简化的误差记录方法, 并发现能以此将其压扩函数简化为一种更为简洁的数值扩展形式。

整数 CDF(2,2)小波由 Cohen, Daubechies 和 Feauveau 在 1992 年提出^[82], 是一种双正交(5,3)小波, 其低通和高通滤波器长度分别为 5 和 3, 都呈现对称性。CDF(2,2)小波经过提升可得到整数到整数双射形式的整数 CDF(2,2)小波, 这种整数形式的小波已被 JPEG2000 标准所采用。实验证明, 同其他整数小波相比, 整数 CDF(2,2)小波应用在可逆水印嵌入时, 可在数据容量和嵌入失真两方面取得比较好的性能指标。设 x_i ($0 \leq i \leq N$) 为一离散整数信号序列, 其正反变换形式如下。

(1) 正变换

1) 样值分组

$$\begin{aligned} s_i &\leftarrow x_{2i} \\ d_i &\leftarrow x_{2i+1} \end{aligned} \quad (5.29)$$

2) 预测

$$d_i \leftarrow d_i - \left\lfloor \frac{1}{2}(s_i + s_{i+1}) + \frac{1}{2} \right\rfloor \quad (5.30)$$

3) 更新

$$s_i \leftarrow s_i + \left\lfloor \frac{1}{4}(d_{i-1} + d_i) + \frac{1}{2} \right\rfloor \quad (5.31)$$

(2) 反变换

1) 逆更新

$$s_i \leftarrow s_i - \left\lfloor \frac{1}{4}(d_{i-1} + d_i) + \frac{1}{2} \right\rfloor \quad (5.32)$$

2) 逆预测

$$d_i \leftarrow d_i + \left\lfloor \frac{1}{2}(s_i + s_{i+1}) + \frac{1}{2} \right\rfloor \quad (5.33)$$

3) 样值合并

$$\begin{aligned} x_{2i} &\leftarrow s_i \\ x_{2i+1} &\leftarrow d_i \end{aligned} \quad (5.34)$$

上面只给出了一维整数 CDF(2,2)的正反变换形式,但注意到整数 CDF(2,2)小波是可分离的,因而可以很容易经过水平和垂直方向两次变换得到二维自然图像的整数小波变换形式。

为了尽量减小比特移位处理带来的大幅值失真,Xuan 等人根据压扩原理设计了一个压缩函数 F ,在此压缩函数中,幅值的比特移位失真可由预先设置的一个阈值 T 进行控制,从而取得数据容量和失真的较佳平衡和整体性能提升。其压缩函数 F 设计为

$$F(x) = \begin{cases} x & |x| < T \\ \text{sign}(x) \cdot \left(\frac{|x| - T}{2} + T \right) & |x| \geq T \end{cases} \quad (5.35)$$

其中 $\text{sign}(x)$ 为 x 的符号,压缩函数、压缩和比特移位操作造成的幅值误差分析如图 5.8 所示。由图 5.8 可知,原始数值 x 经过压缩函数 F 的压缩及左比特移位操作(即幅值乘 2)后,和原始数值的差值 Δx 为

$$\Delta x = \begin{cases} x & |x| < T \\ T & |x| \geq T \end{cases} \quad (5.36)$$

也就是说,无论原始 x 的幅值有多大,经过此压缩函数处理和比特移位后,幅值误差的绝对值总被限制在 T 的范围内。这个处理可以显著降低可逆水印嵌入造成的图像质量失真。压缩函数 F 的数字化版本 $F_Q(x)$ 为

$$F_Q(x) = \begin{cases} x & |x| < T \\ \text{sign}(x) \cdot \left\lfloor \frac{|x| - T}{2} \right\rfloor + T & |x| \geq T \end{cases} \quad (5.37)$$

其中 x 皆取整数。需要注意,由于这里除以 2 后取整, $F_Q(x)$ 不可能为一一映射,因而必然会为扩展函数带来压扩误差,如何有效地记录这个误差也是提高算法性能的关键。

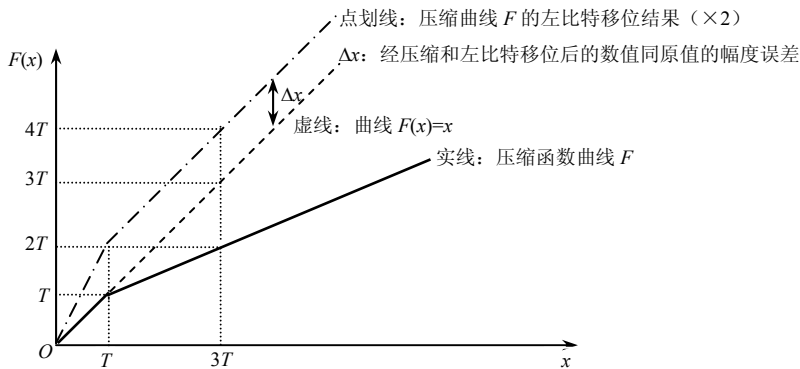


图 5.8 压缩函数 F 及系数幅值误差分析

由式 (5.37) 可知,若 x 保持符号不变, $|x|=T+2k$ 与 $|x|=T+2k+1$, $k \in \mathbb{N}$ 会被压缩为同一数值 $\text{sign}(x) \times (k+T)$, 这样在扩展函数端 E_Q

$$E_Q(x) = \begin{cases} x & |x| < T \\ \text{sign}(x) \cdot (2x - T) & |x| \geq T \end{cases} \quad (5.38)$$

我们将无法分辨 $x = \text{sign}(x) \times (k+T)$ 的原始幅值，只能将此数值恢复为幅值为 $|x| = T+2k$ 的形式，而无法恢复为 $|x| = T+2k+1$ 。即对所有原始幅值为 $|x| = T+2k+1$ 的数值，压扩后都有误差

$$\begin{aligned} q[n] &= x[n] - (E_Q F_Q x)[n] \\ &= \begin{cases} (T+2k) - (T+2k) = 0, & |x| = T+2k \\ (T+2k+1) - (T+2k) = 1, & |x| = T+2k+1 \end{cases} \end{aligned} \quad (5.39)$$

显然，无论 T 取值如何，整数 CDF(2,2) 的高频系数幅值大于或等于 T 的数目总是等于这些高频系数能提供的数据容量（假设每个系数只进行 1 比特左移），同时也等于需要同原始幅值相区分的那些压扩后结果系数的数目。所以记录这些用以区分原始幅值的压扩误差一个比较好的方法是用“0”“1”比特区分每个压扩后系数的误差，并将此比特串存储在相应每个压缩和左比特移位后空出的 LSB 位上。下面图 5.9 的例子说明了这个过程，其中 $|x|$ 为原始幅值， $|F_Q(x)|$ 为经压缩函数 F_Q 压缩后的幅值， $|2F_Q(x)|$ 为经过 1 比特左移位后的幅值， $|2F_Q(x)'|$ 为含有区分原始幅值信息的幅值， A 为用于区分原始幅值的比特串， $|E_Q F_Q(x)|$ 为扩展函数作用后的幅值， $T=3$ 为这个例子中的压扩函数用到的幅值阈值。

$ x $:	1	2	3	4	5	6	7	8	9	10	...
$ F_Q(x) $:	1	2	3	3	4	4	5	5	6	6	...
$ 2F_Q(x) $:	2	4	6	6	8	8	10	10	12	12	...
$ 2F_Q(x)' $:	2	4	6	7	8	9	10	11	12	13	...
A :			0	1	0	1	0	1	0	1	...
$ E_Q F_Q(x) $:	1	2	3	3	5	5	7	7	9	9	...

图 5.9 数值压扩方法示例

由图 5.9 可知，压扩误差 $|x| - |E_Q F_Q(x)|$ 甚至直接等于原始幅值区分信息 A ，这是因为由式 (5.37) 和式 (5.38) 决定的压扩误差正好也是 0、1 交错的情况，即式 (5.39)。在此例中，将压扩误差为 0 的情况用“0”来表示，将压扩误差为 1 的情况用“1”来表示，所以原始幅值区分信息 A 可被直接用来表示压扩误差。当然也可以换一种表示方法，即用“0”表示压扩误差 1，而用“1”表示压扩误差 0，但实验证明，该表示法下的 $|2F_Q(x)|$ 同原始幅值 $|x|$ 之间的失真会变得更大。分析可知，前一种表示方法下 $|2F_Q(x)'|$ 同原始幅值 $|x|$ 之间的误差在 $|x| \geq T$ 时均为 T （图 5.9），而后一种表示方法下，相应的误差为 $T+1$ 和 $T-1$ 。由于失真计算是幅值误差的平方关系，在原始相邻幅值分布较为均衡的情况下，显然前一种表示方法的总体失真更小。

由上面的分析还能发现，如果采用第一种表示方法，即将压扩误差为 0 的情况用“0”来表示，将压扩误差为 1 的情况用“1”来表示，就总能有如下关系

$$|2F_Q(x)'| = \begin{cases} 2 \cdot |x|, & |x| < T \\ |x| + T, & |x| \geq T \end{cases} \quad (5.40)$$

由此可以得到数值扩展水印嵌入的一个简化形式

$$S = \begin{cases} 2C + W, & |C| < T \\ C + \text{sign}(X) \cdot T, & |C| \geq T \end{cases} \quad (5.41)$$

易知式 (5.41) 在幅值失真方面等效于式 (5.40)，该数值扩展水印嵌入方法和 Xuan 等人的压扩方法在效果上近似，但形式上更简明，计算量更小，称为阈值内数值扩展方法。

5.5.3 整数变换域直方图移位技术

杨等人^[83]将空间域直方图移位技术应用到整数 DCT 域内，利用整数 DCT 系数靠近 0 值集中分布的特点，可以显著提高直方图移位的效率，提升可逆水印算法的整体性能。具体的直方图移位技术已经在前面做过介绍，杨等人把它用在 8×8 整数 DCT 域内，并提出一种将附加信息存储在空间域指定像素块的 LSB 位的方案，能够替代前面 5.5.1 节介绍的块甄别方法以提高算法的整体运行效率。

这个算法首先需要在整数 DCT 域产生系数的直方图，之后使用前面介绍过的直方图移位技术产生空余的比特空间以嵌入水印信息。假设 C 为一幅 256 灰度的图像，首先将其划分为 B 个 8×8 的像素块 $C_i (i=1, 2, \dots, B)$ ，如图 5.10 所示。然后经过系数分组可得到 63 个 AC 系数分组，这里标记为 $G(p, q)$ ($0 \leq p, q \leq 7$ ，且 $p+q > 0$)，对每个系数分组可统计出自己的直方图 $H(p, q)$ ($0 \leq p, q \leq 7$ ，且 $p+q > 0$)。从直方图 $H(p, q)$ 得到相应的绝对值分布直方图 $H_{abs}(p, q)$ ，即将原来 $H(p, q)$ 的负刻度（小于 0）部分按镜像融合到原来正刻度（大于 0）部分。依然利用系数分组直方图围绕 0 值集中分布的特性，对直方图进行移位处理并嵌入水印。由于直方图移位操作是向右（远离 0 刻度的方向）进行，只会在原来数值的基础上增加其幅值而不会改变原有幅值的符号，因而能够直接在绝对值直方图 $H_{abs}(p, q)$ 上进行移位，这样要比在 $H(p, q)$ 上选择峰点和零点进行移位要有效得多。图 5.11 给出了在 $H_{abs}(p, q)$ 上进行移位嵌入水印的示意图，直方图移位可逆水印的具体原理和过程已在前面详细介绍过，这里不再赘述。

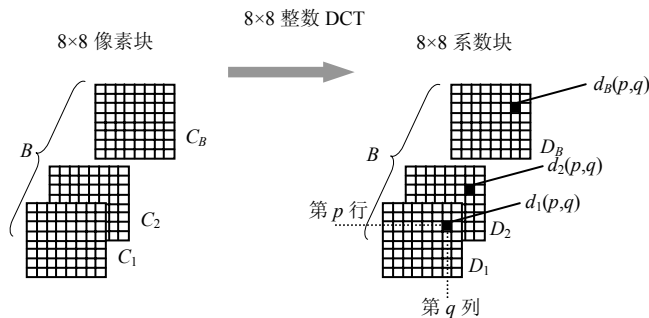


图 5.10 8×8 整数 DCT 系数分组

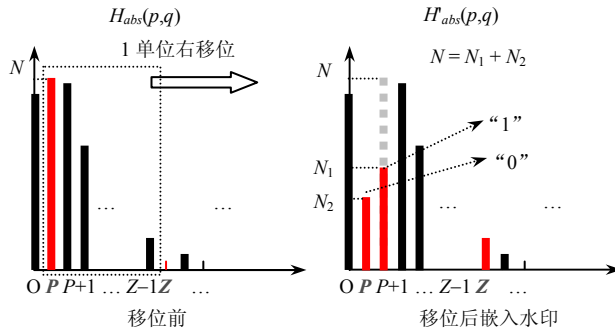


图 5.11 8×8 整数 DCT 域分组系数绝对值直方图移位可逆水印嵌入

从实验中发现,不同图像的系数分组直方图中峰点 P 的位置可能不一样,出现在绝对值刻度为 2 或者 3 的峰点比较多。也就是说,对不同的系数分组直方图,都需要记录其峰点 P 和零点 Z ,作为附加信息一起嵌入到载体图像中去,将这一类附加信息记为 OH_{PZ} 。另一类需要同时嵌入载体图像的是块分类信息,类似 5.5.1 节中判别像素块是否适合水印嵌入的方法,通过系数修改对空间域像素值产生误差的估计,在空间域内设定阈值,从而将原始图像像素块分为适合水印嵌入和不适合水印嵌入两类,此分类信息同样用附加信息记录,记为 OH_L 。

如何在载体图像中记录 OH_{PZ} 和 OH_L 也是需要考虑的问题。5.5.1 节采用了块甄别方案,实现了无需先验信息便可直接在载体图像中提取附加信息。同样此处也可以采用这种块甄别方法。但这种方法也有其缺陷:一是计算过程复杂,计算量大;二是附加信息的长度受限於分类Ⅲ所能提供的比特空间。对于细节丰富的图像而言,分类Ⅲ的像素块数目可能会很少,反而不适合水印嵌入的像素块(分类Ⅰ)数目会很多,这样用于块分类的附加信息可能会很长,甚至超出分类Ⅲ所能提供的比特空间而使算法不可用。为解决这个问题,杨等人在整数 DCT 域直方图移位算法^[83]中提出了一种使用密钥 Key 来预先选择像素块进行附加信息嵌入的方法,如图 5.12 所示。首先在原始图像空间域使用密钥 Key 选择 K 个 8×8 像素块,再对除这 K 个像素块外的所有像素块作 8×8 整数 DCT 变换,然后对系数分组作直方图移位,估计系数改变给像素值带来的误差,并对原始像素块进行分类,找出那些对像素溢出敏感而在算法中不会被改动的像素块。注意,此时原有像素块分成了三部分,第一部分 A_1 是由密钥 Key 所选定,用于嵌入附加信息;第二部分 A_2 是由误差估计方法找出的不会被水印算法改动的像素块;第三部分 A_3 是直方图移位和水印嵌入的对象。

在算法中,先由密钥 Key 选定 A_1 ;然后计算出剩余像素块(A_2 和 A_3)组成的分组系数直方图和所有的峰点、零点值信息 OH_{PZ} ;之后由误差估计方法找出 A_2 ,记为附加信息 OH_L ;再重新以 A_3 为基础计算新的直方图,仍沿用原有 OH_{PZ} 信息进行直方图移位;最后在 A_1 中像素值的 LSB 位上记录 OH_{PZ} 和 OH_L 信息,而 A_1 中像素值的原始 LSB 位信息连同水印数据嵌入到 A_3 所能提供的比特空间中去。在水印提取和图像恢复阶段,首先根据密钥 Key 提取 A_1 的位置;由 A_1 中像素的 LSB 位可提取 OH_L 信息以获得 A_3 的位置,提取 OH_{PZ} 信息以获得峰点和零点信息从而对 A_3 进行水印提取;从 A_3 中提取 A_1 中像素原始 LSB 位以恢复 A_1 内所有像素块,并从 A_3 中提取水印数据后恢复所有 A_3 内的像素块。

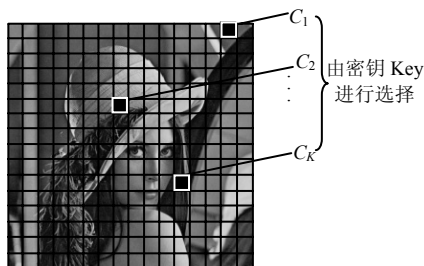


图 5.12 密钥控制选择像素块进行附加信息嵌入

整数 DCT 域直方图移位可逆水印算法引入的失真小,特别适合于要求高质量含水印图像的场所。它的高 PSNR 区域的数据容量要显著高于前面介绍的整数 DCT 域比特移位方法,而与小波域压扩比特移位方法性能接近。但是实验表明,如果在同样的 CDF(2,2) 整数小波域使用直方图移位方法,其性能要高于 Xuan 等的压扩比特移位方法。

5.6 压缩域无损信息隐藏技术

5.6.1 压缩域无损信息隐藏技术和应用要求

在日常应用中,绝大多数图像数据都以各种压缩编码形式存在于数字存储设备上。对一些有损压缩算法如 JPEG 和矢量量化 (Vector Quantization, VQ) 而言,在原始图像中嵌入的秘密信息很可能会在压缩过程中经历一定程度的损失。对于版权保护等适合鲁棒水印的应用场合,有损压缩过程对水印算法的鲁棒性有相当高的要求;对适合脆弱水印、签名嵌入和可逆水印等同精确认证相关的应用而言,有损压缩会完全破坏水印的功能,使得图像内容的水印认证根本无法在像素值上进行。另一方面,若水印信息是嵌在图像未压缩之前的像素值上,那么在从图像的压缩格式提取水印前,必须先对图像进行解压使其恢复到像素值的状态然后才能进行水印提取。这一过程的执行效率同图像压缩编码和解码的计算复杂度紧密相关,这样的水印算法可能无法满足某些实时应用的要求。

基于以上两方面的原因,压缩域信息隐藏技术被提出且用来直接对图像的有损压缩格式数据(或称码流)进行秘密信息嵌入,这样不仅使秘密信息免于压缩过程的干扰,而且部分甚至完全省去了解码和重新编码过程。目前的压缩域信息隐藏技术主要针对 JPEG 和 JPEG2000 格式,也有一些针对 VQ 压缩算法和 BTC (Block Truncation Coding) 压缩算法。事实上,一部分算法只是选取较为鲁棒的特征或者将压缩过程当作秘密信息嵌入的一个工具,本身仍需要参与全部甚至部分解码和编码过程,因而严格地说只能算是针对某种压缩算法的信息隐藏,但是其中很多特征提取和水印嵌入的思想可以直接移植到压缩域中,这里从技术上也将其分为压缩域信息隐藏这一类。

压缩域无损信息隐藏算法最早出现在 Fridrich 的文献[84]中,用于在 JPEG 格式的图像码流上对图像内容作可逆认证。Fridrich 还进一步探讨了在不引起 JPEG 格式文件长度变化的情况下在码流上进行图像内容的可逆认证。近些年来,本书作者一直关注矢量量化域和 BTC 域的无损信息隐藏技术。总地说来,压缩域无损信息隐藏算法的研究方兴未艾,而且往往直接面向应用。

当从压缩域的角度考虑无损信息隐藏算法时,所关心的对象实际上已从原始图像的像素值形式转为压缩域的码流形式。这种嵌入对象的转变也使无损信息隐藏嵌入算法所对应的技术和应用要求有所转变,下面首先给出有损压缩算法和压缩域无损信息隐藏算法的基本框架(图 5.13),然后在此基础上讨论压缩域无损信息隐藏算法需要具备的基本性质,以及不同应用中人们希望压缩域无损信息隐藏算法还能兼有的一些附加性质。

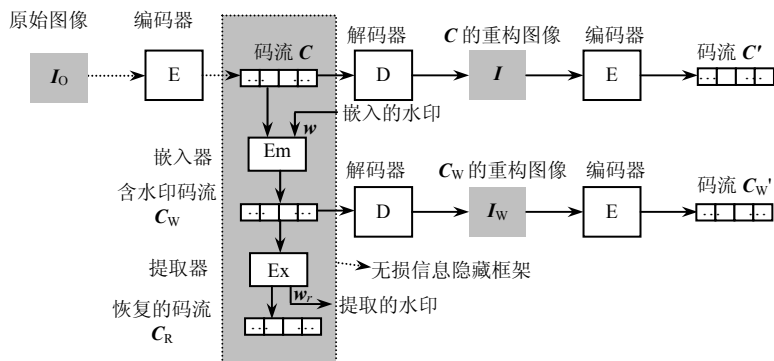


图 5.13 有损压缩和压缩域无损信息隐藏框架

在图 5.13 中, I_0 为原始图像, E 为具有量化器的有损编码器, C 为 E 的输出码流, 也是载体对象。 D 是同 E 对应的解码器, E_m 和 E_x 分别为秘密信息嵌入器和提取器。可以看出, 由于 E 的量化作用, 从 I_0 到 C 实际上是一个多对一的映射, 如果直接在 I_0 上嵌入水印 w , 那么量化过程可能造成秘密信息的损失, 以致整个嵌入过程不再可逆。当把载体对象由 I_0 转变为 C 时, 相应也需要规定一些压缩域无损信息隐藏算法应具备的基本性质和一些应用下人们希望其具备的一些附加性质。

压缩域无损信息隐藏算法所应具备的基本性质如下所示。

- (1) 可逆性: C 和 w 的完全恢复, 即 $C = C_R$ 且 $w = w_r$;
- (2) 不可区分性: C 中的 w 在数据格式上和 C 不可分隔, 并且在隐写分析意义上与 C 的主体部分不可区分。这保证解码器 D 对 C_w 的解码可以像对 C 的解码一样顺畅, 无法意识到 w 的存在;
- (3) 保真度: w 的嵌入不会造成重构图像 I_w 和原始图像 I_0 之间较大的失真;
- (4) 图像的像素域尺寸不变: 重构图像 I_w 和原始图像 I_0 的像素域尺寸保持不变。

除了以上这些压缩域无损信息隐藏算法必须具备的基本性质外, 在一些应用下还希望它具有有一些附加性质。

(1) 兼容性: 无损信息隐藏算法应当同原有的有损压缩算法尽量兼容, 对原有编解码器的改动越少越好;

(2) 码流的尺寸不变性: 在一些应用场合, 特别是存储受限或者隐写性要求较高的场合, 希望秘密信息嵌入后, 码流长度基本保持不变;

(3) 在解码图像中秘密信息的继承性: 有时, 无损信息隐藏算法无法保证秘密信息 w 从 C_w 到 C'_w 的继承性, 例如在解码算法中可能存在防止像素值溢出的硬截断处理, 使得秘密信息在这个过程中遭到损失, 而在某些场合人们可能希望还能从 I_w 通过编码得到的 C'_w 中找到秘密信息;

(4) 编码器和秘密信息嵌入器分离: 这样会方便二者各自的运行分析和结构改进。

下面介绍针对块截断编码和 JPEG2000 这两个压缩域的一些典型无损信息隐藏技术。

5.6.2 BTC 压缩域无损信息隐藏

1. 块截断编码 (BTC)

块截断编码是由 Delp 和 Mitchell 于 1979 年发明的。它是一种非常简便有效的图像压缩方法, 在压缩率和视觉失真程度方面都有着卓越的表现。它的核心思想是先将一幅灰度图像分割成若干个不相重叠的区块, 然后对每一个区块实行某种已知的矩保留量化 (Moment Preserving Quantizer, MPQ) 方法, 将相应的区块量化成一对高低量化值以及一个与原区块大小相同的位平面 (Bit-plane) 来表征量化判决结果。Delp 和 Mitchell 提出这一编码方法时, 在矩保留量化部分不仅保留了原始数据的一阶矩 (即平均值), 还保留了原始数据的二阶矩。自块截断编码于 1979 年被提出之后直到目前, 在这三十多年间不断地有人对这一编码方式进行改进, 由此可见研究人员对 BTC 压缩的研究热度。下面介绍目前最为常用的由 Lema 和 Mitchell 于 1984 年改进的绝对矩块截断编码 (AMBTC) 方法。

对于大小为 $M \times N$ 的灰度图像, AMBTC 编码器首先将其分成 $n \times n$ 大小的不重叠子块, 一般来说, n 值为 4。对于每个子块, 首先计算像素均值 \bar{x} , 再以此均值将子块分为两部分: 大于等于均值的一类, 以及小于均值的另一类。假设 $x_{u,v}$ ($u, v = 1, 2, \dots, n$) 是

子块在点 (u, v) 上的像素值，二进制位图 B 用于记录这两类像素，大于等于均值的一类用“1”标识，小于均值的另一类用“0”标识

$$B(u, v) = \begin{cases} 1 & x_{u,v} \geq \bar{x} \\ 0 & \text{其他} \end{cases} \quad (5.42)$$

$B(u, v)$ 表示位图 B 在位置 (u, v) 的比特值。在该子块中，计算大于等于均值的一类像素平均值，记为高均值 h ；计算另一类像素平均值，记为低均值 l

$$l = \begin{cases} \frac{1}{N_L} \sum_{x_{u,v} < \bar{x}} x_{u,v} & N_L \neq 0 \\ \bar{x} & \text{其他} \end{cases} \quad (5.43)$$

$$h = \frac{1}{n \times n - N_L} \sum_{x_{u,v} \geq \bar{x}} x_{u,v} \quad (5.44)$$

其中， N_L 表示小于均值的像素个数。 l 、 h 和 B 组成每个子块的编码，以 (l, h, B) 形式存储。

解码 AMBTC 压缩图像时，首先从码流中读取各子块压缩码 (l, h, B) 。接着扫描位图 B ，比特值为 0 则该子块对应位置像素值重建为低均值 l ；比特值为 1 则重建为高均值 h 。图 5.14 是 AMBTC 编码的一个示例。图 5.14 (a) 是一原始图像子块，求得均值为 129.6。据此均值，原始图像子块被分成两部分，如图 5.14 (b) 所示，大于等于均值的用“1”标示，其对应区域像素均值 $h = 168$ ；小于均值的像素用“0”标示，其对应区域像素均值 $l = 106$ 。由子块压缩码 (l, h, B) 可重建图像子块，如图 5.14 (c) 所示。

AMBTC 的好处在于四个方面，即它非常快，它易于实现，它有较低的计算要求，它保护了重构图像的质量并保留了边缘。很明显，对于每个 256 灰度等级的 4×4 大小的图像块，低均值和高均值分别用 8 比特编码，位平面需要 16 比特，所以原始 AMBTC 的比特率是 $(8+8+16)/16=2\text{bpp}$ 。

2. 基于 AMBTC 位平面翻转的无损信息隐藏

在上面所叙述的三元组 (l, h, B) 中，如果将 l 和 h 的次序交换，那么在图像重建的过程中只需要将 B 中所有的比特位翻转一下即可保持原有压缩图像不变。Hong 等人正是利用这种思想，提出了一种基于 BTC 域灰度图像的无损数据隐藏方法^[85]。Hong 等人的方法大致如下。

196	190	157	113
186	104	111	114
131	105	105	112
90	102	149	108

(a) 原始图像子块

1	1	1	0
1	0	0	0
1	0	0	0
0	0	1	0

(b) 位图 B

168	168	168	106
168	106	106	106
168	106	106	106
106	106	168	106

(c) 重建图像子块

图 5.14 AMBTC 编码示例

先将一幅给定的灰度图像进行 AMBTC 编码，每一数据块唯一对应一个三元组 (l, h, B) ，如果 $l \neq h$ 则称该数据块为可嵌入，如果 $l = h$ 则称该数据块为不可嵌入。

Hong 等人将隐藏信息嵌入到所有可嵌入的数据块中。对于每一个可嵌入的数据块，如果要嵌入的隐秘信息为“1”，那么交换一下 l 和 h 的次序，同时把对应的 B 中的所有

比特位翻转；如果要嵌入的隐秘信息为“0”，那么该数据块维持不变。

由于在 AMBTC 编码过程中 $h \geq l$ ，因而在没有嵌入数据之前，所有可嵌入的数据块中均满足： $h > l$ 。在抽取隐秘信息的过程中，如果发现数据块中 l, h 的次序是正常的，即数值较小的在前，数值较大的在后，那么该数据块嵌入的数据是“0”。当发现数据块中 l, h 的次序是数值较大在前，数值较小的在后，即 (h, l, B) 时，就说明该数据块已嵌入数据“1”，提取出隐秘数据“1”，将 l, h 的次序恢复到正常的 (l, h, B) ，同时将 B 中所有比特位进行翻转。

经过这样的抽取过程，不仅隐秘数据可以完整地得到，而且原始的 AMBTC 压缩域图像也可以无失真的获得。这就实现了无损数据隐藏的目标。

Hong 等的方法有效地利用了 AMBTC 编码方式中高低平均值的次序信息来进行无损数据隐藏。但该方法有一个弊病，就是当高低平均值相等时，该算法无法对其进行嵌入处理，这无疑是浪费了一定的嵌入空间。

通过观察发现在 AMBTC 中，如果高低平均值相同，那么整个 B 中比特数据实际上是没有任何作用的。Chen 等人^[86]正是看到了这一点，提出了一种对 Hong 等人的方法进行改进的新方法：如果在三元组 (l, h, B) 中 $l = h$ ，那么整个 B 都可以作为嵌入空间进行数据隐藏。由此，当 AMBTC 压缩域图像中出现高低平均值相等的数据块时，Chen 等人的方法在隐秘数据的嵌入容量上就会大大增加。如果 AMBTC 压缩域图像不存在高低平均值相等的数据块，那么 Chen 等人的方法就蜕化为 Hong 等人的方法。

3. 基于直方图移位的 AMBTC 域无损信息隐藏

上面描述的 Chen 等人的无损数据隐藏方法主要是针对由 AMBTC 编码生成的高低平均值进行处理，算法改变的只是高低平均值的相对位置，并没有对数据本身进行修改。更进一步地讲，在正常的 BTC 压缩数据中，整个图像的高低平均值的次序都是固定的，而 Chen 等人的方法改变了这一规则，因而这种数据隐藏的方法容易被察觉，非法人员可以通过统计高低平均值次序的变化轻易地得到原始的隐藏数据。本书作者以此为出发点，在 Chen 等人工作的基础上引入直方图移位技术，使得不仅高低平均值的次序发生变化而且高低平均值的数值本身也进行无损地修改^[87]。这样一来如果非法人员仍然通过统计高低平均值次序的变化来窥探数据，那么他会改变直方图移位技术所依赖的数据，造成由直方图移位所隐藏的数据无法恢复。

该算法先对一幅图像进行 AMBTC 编码，然后对两列高低平均值数据分别进行直方图移位操作，实现第一级的无损数据隐藏，如图 5.15 和图 5.16 所示。最后应用上面提到的 Chen 等人的算法进行第二级的无损数据隐藏。算法具体描述如下。

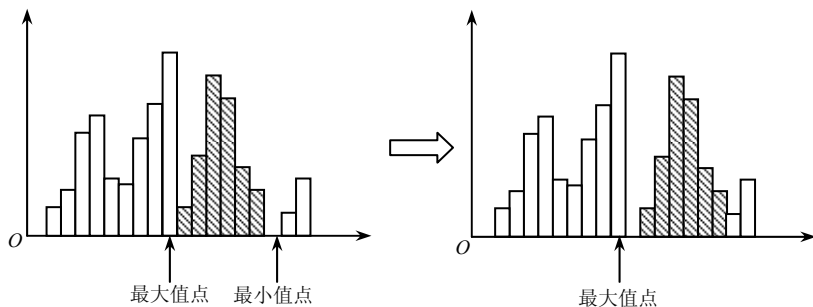


图 5.15 直方图移位示意图

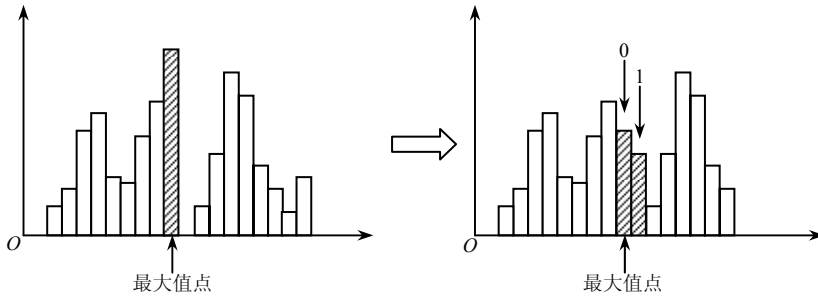


图 5.16 数据嵌入示意图

(1) 数据嵌入部分

1) AMBTC 编码: 给定一幅大小为 $M \times N$ 的 256 阶灰度图像 C , 取块大小为 $m \times n$, 将其进行 AMBTC 编码, 得到高低平均值序列分别为 $L = \{l_i | 1 \leq i \leq (M \times N)/(m \times n)\}$, $H = \{h_i | 1 \leq i \leq (M \times N)/(m \times n)\}$, 以及位平面 $B = \{B_i | 1 \leq i \leq (M \times N)/(m \times n)\}$ 。其中 B_i 的大小为 $m \times n$, 且与 l_i, h_i 构成三元组 (l_i, h_i, B_i) 。

2) 对高平均值序列进行直方图移位, 步骤如下。

① 计算高平均值序列 H 的直方图: $H_h(x), x \in [0, 255]$ 。

② 在 $H_h(x)$ 中找到最大值点 x_{\max} 和最小值点 x_{\min} , 并且记录下来。此处不妨设 $x_{\max} < x_{\min}$ (对 $x_{\max} > x_{\min}$ 的情况的处理, 移位方向相反即可), 如图 5.16 所示。若 $H_h(x_{\min}) \neq 0$, 记录所有值等于 x_{\min} 的像素点的位置。

③ 移位 (如图 5.15): 若 $x_{\max} < x_{\min}$, 遍历整个序列 H , 当 $h_i \in [x_{\max} + 1, x_{\min})$ 时, 将 h_i 的值加 1; 若 $x_{\max} > x_{\min}$, 遍历整个序列 H , 当 $h_i \in (x_{\min}, x_{\max} - 1]$ 时, 将 h_i 的值加 1。

④ 嵌入数据 (图 5.16): 再次遍历整个序列 H , 当 $h_i = x_{\max}$ 时, 若要嵌入的隐藏数据为 '1', 则将 h_i 的值加 1; 若要嵌入的隐藏数据的值为 '0', 维持 h_i 不动。

3) 对低平均值序列 L 进行直方图移位: 按照类似 2.1) 至 2.4) 的步骤对低平均值序列 L 进行直方图移位。

4) 修改高低平均值的相对次序: 同时遍历经过直方图移位后的高低平均值序列 H 和 L , 执行下列规则。在 $h_i' \neq l_i'$ 的情况下, 如果待嵌入的数据为 '1', 则将 l_i' 存入到 h_i' 在高平均值序列中的位置, 将 h_i' 存入到 l_i' 在低平均值序列中的位置, 即交换 h_i' 和 l_i' 。同时将与 h_i' 和 l_i' 对应的 B_i 中所有比特位翻转; 如果待嵌入的数据为 '0', 维持 (l_i', h_i', B_i) 不动。在 $h_i' = l_i'$ 的情况下, 直接将 $m \times n$ 位的待嵌入数据替换 B_i 。

(2) 数据提取和图像恢复部分

对数据提取的操作大致是数据嵌入的逆过程。

1) 统计高低平均值的相对次序: 同时遍历含有隐藏数据的高低平均值序列 H^W 和 L^W , 执行下列操作。

在 $h_i^W \neq l_i^W$ 的情况下, 如果 $h_i^W < l_i^W$, 提取出数据 '1', 再将 l_i^W 存入到 h_i^W 在高平均值序列中的位置, 将 h_i^W 存入到 l_i^W 在低平均值序列中的位置, 即交换 h_i^W 和 l_i^W 。最后将与 h_i^W 和 l_i^W 对应的 B_i^W 中所有比特位翻转; 如果 $h_i^W > l_i^W$, 提取出数据 '0', 然后维持 (l_i^W, h_i^W, B_i^W) 不动。

在 $h_i^W = l_i^W$ 的情况下, 直接从 B_i^W 提取出 $m \times n$ 位的隐藏数据。

2) 对低平均值进行直方图移位的逆过程如下。

① 提取数据: 按顺序遍历经过 1) 处理过后的低平均值序列 L'' 。若 $x_{\max} > x_{\min}$, 则

当 $l_i'' = x_{\max} + 1$ 时, 提取出数据 ‘1’, 然后将 l_i'' 减 1; 若 $l_i'' = x_{\max}$, 则提取出数据 ‘0’。若 $x_{\max} < x_{\min}$, 则当 $l_i'' = x_{\max} - 1$ 时, 提取出数据 ‘1’, 然后将 l_i'' 减 1; 若 $l_i'' = x_{\max}$, 则提取出数据 ‘0’。

② 恢复: 再次按顺序遍历低平均值序列 L'' 。若 $x_{\max} > x_{\min}$, 则当 $l_i'' \in (x_{\max} + 1, x_{\min})$, 将 l_i'' 减 1。若 $x_{\max} < x_{\min}$, 则当 $l_i'' \in [x_{\min}, x_{\max} - 1)$, 将 l_i'' 减 1。如果 $H_l(x_{\min}) \neq 0$ (前已记录所有值等于 x_{\min} 的位置) 再根据记录把相应的位置的像素值置为 x_{\min} 。

3) 对低平均值进行直方图移位逆过程: 按照类似步骤 2) 对高平均值 H'' 进行操作。

4. 实验结果

实验中, 采用 “Lena”、“Pepper”、“Bridge”、“Boat”、“Goldhill” 和 “Jet_F16” 这 6 幅大小同为 512×512 的 256 阶灰度图像作测试图像, 如图 5.17 所示。在 AMBTC 编码过程中时, 块的大小选为 4×4 , 编码得到高低平均值构成的图像如图 5.18 所示, 位平面数据如图 5.19 所示, 压缩后的图像效果如图 5.20 所示, 而嵌入信息后的图像效果如图 5.21 所示。失真度评价使用峰值信噪比 PSNR。实验结果见表 5.2。可以看出, 应用直方图移位技术可使得嵌入容量较 Chen 等人的算法得到提升。对于 “Lena”、“Pepper”、“Bridge” 和 “Jet_F16”, Chen 等人的算法的嵌入容量比 16384 ($=128 \times 128$) 要大, 这是因为在这四幅图像中都存在一些高低平均值相等的块, 因而此刻 Chen 等人的方法要比 Hong 等人的方法在性能上较优。表 5.2 表明了在提取出秘密信息后, 图像的 PSNR 与嵌入信息之前一样, 这就说明在提取信息之后图像得到了完整地恢复, 从而证明 Chen 方法结合直方图移位技术满足无损信息隐藏要求。



图 5.17 六幅测试图像

表 5.2 算法的性能比较

图 像	Lena	Pepper	Bridge	Boat	Goldhill	Jet_F16
AMBTC 的 PSNR(dB)	32.041	31.595	28.585	31.151	33.163	31.033
结合直方图移位的算法的嵌入容量(Bits)	16762	212927	17567	17175	16798	17494
Chen 等人的算法的嵌入容量(Bits)	16414	20659	17194	16384	16384	16444
高低平均值相等块的数量	2	285	49	0	0	7
结合直方图移位的算法的 PSNR(dB)	32.031	31.583	28.572	31.146	33.154	31.029
Chen 等人算法的 PSNR(dB)	32.041	31.595	28.585	31.151	33.163	31.033
结合直方图移位的算法恢复图像的 PSNR(dB)	32.041	31.595	28.585	31.151	33.154	31.033

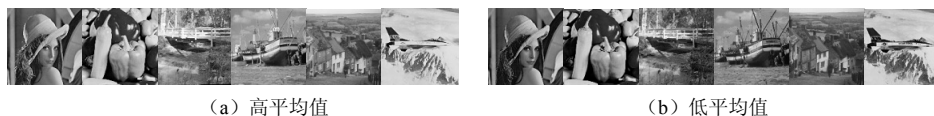


图 5.18 由 AMBTC 编码生成的高、低平均值序列

5.6.3 JPEG 压缩域无损信息隐藏

JPEG 压缩域的无损信息隐藏逐步成为学者们关注焦点，因为 JPEG 格式是大多数图片所采用的压缩格式。针对 JPEG 压缩的特点，进行无损信息隐藏可以修改的位置包括量化 DCT 系数、JPEG 码流中的**可变长整数**（Variable Length Integer, VLI）、量化表等，并可以将空间域的技术，包括无损压缩替换、直方图移位和差值扩展等借鉴到 JPEG 压缩域中来。一些典型的算法思想如下：Fridrich 等人^[84]最早提出对量化 DCT 系数进行无损压缩，从而得到嵌入空间进行数据嵌入；同时提出一种基于修改量化表和数值扩展的操作。Chang 等人^[88]通过修改 DCT 块零值边界系数，将信息嵌入到两连续零值系数中，达到较高的嵌入容量。Xuan 等人^[89]提出一种基于直方图对的 JPEG 图像无损数据隐藏方法，通过调整直方图在量化 DCT 中频系数中嵌入隐藏数据。Fridrich 等人在文献[90]中首次提出了保持载体文件大小不变的无损信息隐藏算法，利用 JPEG 压缩过程中范式 Huffman 编码原则，将秘密信息嵌入到 JPEG 码流中的 VLI，保持其码流长度不变。受 Fridrich 等人的启发，本书作者提出联合修改量化表和 DCT 系数的 JPEG 图像无损信息隐藏方案^[91]，首先分析修改每个量化 DCT 系数对图像质量造成的影响，选取一个合适的嵌入顺序来保证嵌入后载体的图像质量并避免文件大小的过快增长；之后通过降低某些量化表元素，同时提升相应量化表系数产生冗余空间以进行信息嵌入，联合修改中辅以一个调整量来尽可能保持图像质量。下面先简要介绍 Fridrich 等人的方法，然后介绍本书作者的一种算法。



图 5.19 由 AMBTC 编码生成的位平面

1. Fridrich 等人的方法

在文献[84]中，Fridrich 等人提出两种针对 JPEG 图像的可逆水印技术，第一种方法对量化 DCT 系数进行无损压缩，从而将水印嵌入到压缩节省出来的空间；而第二种方法采用基于修改量化表和数值扩展的方法。其中，第二种方法是根据量化表元素 $Q(u, v)$ 的奇偶性来修改以嵌入水印。如果 $Q(u, v)$ 是偶数，则将其除以 2，并且对应的所有量化 DCT 系数 $D(u, v)$ 都乘以 2；而如果 $Q(u, v)$ 是奇数，则将其修改为 $\lfloor Q(u, v) / 2 \rfloor$ ，对应的所

有量化 DCT 系数 $D(u, v)$ 都乘以 2。之后, 在将二进制信息嵌入修改后量化 DCT 系数的 LSB 中。为了保证该方法是可逆的, 需要将 $Q(u, v)$ 为奇数的信息一同嵌入到图像中。

2. 联合修改量化表和 DCT 系数的无损信息隐藏方案

Fridrich 等人的方法给了我们很好的启示, 本书作者对该方法进行扩展, 提出基于模操作的 k 进制位大容量无损信息隐藏算法^[91], 通过减小某些量化表元素、提升对应量化 DCT 系数以产生冗余空间进行信息嵌入, 同时联合修改中辅以调整量并选取适当嵌入位置以得到更好的图像质量。



图 5.20 经过 AMBTC 压缩过的图像



图 5.21 嵌入秘密信息后的图像

(1) 嵌入位置的选取

为了使因嵌入信息对图像质量所造成的影响尽可能的小, 对于嵌入位置的选取需要仔细分析。在 DCT 变换域, 以往的信息隐藏技术倾向于优先选择中频 DCT 系数进行修改以嵌入信息。对于 JPEG 图像, 由于量化步骤的引入, JPEG 图像文件保存的是量化后的 DCT 系数, 这时情况就不同于原始的 DCT 域。因此, 先前的技术研究可能会影响到在 JPEG 域的嵌入位置选择。为了更详细的测试每个量化 DCT 系数对图像质量的影响, 我们做了理论上的分析以及实验研究, 这里采用 PSNR 来进行评估。

因为存储在 JPEG 图像文件的是量化表及量化 DCT 系数，所以需要通过计算重构 DCT 系数 $\tilde{F}(u, v) = D(u, v) \times Q(u, v)$ 。当我们在量化 DCT 系数 $D(u, v)$ 上加一个数 a ，就相当于 $\tilde{F}(u, v)$ 加上了 a 乘以对应的量化表元素 $D(u, v)$ 。从第 2 章式 (2.28) 的 JPEG 标准量化表可以看出，不同位置的量化表元素各不相同。一般低频元素小于中频元素，而中频元素又小于高频元素。例如，当 $D(i, j)$ 加上 2 时，有 $F^w(i, j) = \tilde{F}(u, v) + 2Q(i, j)$ 。由第 2 章给出的 DCT 逆变换公式 (2.26) 可知，对于当 $u \neq i$ 或 $v \neq j$ 都有差值 $\Delta F(u, v) = F^w(u, v) - \tilde{F}(u, v) = 0$ 的情况时，则差值 $\Delta f(x, y) = f^w(x, y) - \tilde{f}(x, y)$ 由 $\Delta F(i, j)$ 唯一确定。基于这点，我们可对每个量化 DCT 系数对图像质量的影响进行测试。

在一个 8×8 分块中，若在 $D(i, j)$ 上加 1 并保持其他 $D(u, v)$ 不变，则引入差值如下

$$\begin{aligned} \Delta f(x, y) &= f^w(x, y) - \tilde{f}(x, y) \\ &= \frac{1}{4} \left[\sum_{u=0}^7 \sum_{v=0}^7 C(u)C(v)\Delta F(u, v) \cos \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16} \right] \\ &= \frac{1}{4} C(i)C(j)Q(i, j) \cos \frac{(2x+1)i\pi}{16} \cos \frac{(2y+1)j\pi}{16} \end{aligned} \quad (5.45)$$

根据 $\Delta f(x, y)$ 很容易计算出对应的理论上的 PSNR 值。实际测试是在六幅 512×512 大小的 256 阶灰度图像上进行，信息隐藏采用对每个分块 $D(i, j)$ 的 LSB 取反的方法，测试结果如图 5.22 所示。图中，PSNR 是由原始 JPEG 图像和伪装 JPEG 图像计算得出，DCT 系数位置按 zigzag 方式排序，蓝色带圈实线为理论计算结果，其余彩色实线是六张图像实际测试的结果。从图 5.22 可以看出实际测试结果几乎一样，相互重叠。理论计算结果与实际测试结果大致一样，细微差别主要是由于实际的 JPEG 算法中 FDCT 和 IDCT 采用快速实现算法，而不是数学定义。图 5.23 是原始未压缩图像和伪装 JPEG 图像计算得出的 PSNR 结果。六张原始 JPEG 图像与其原始未压缩图像间的 PSNR 值分别为 27.49dB、33.50dB、35.40dB、33.80dB、34.44dB 和 33.55dB。

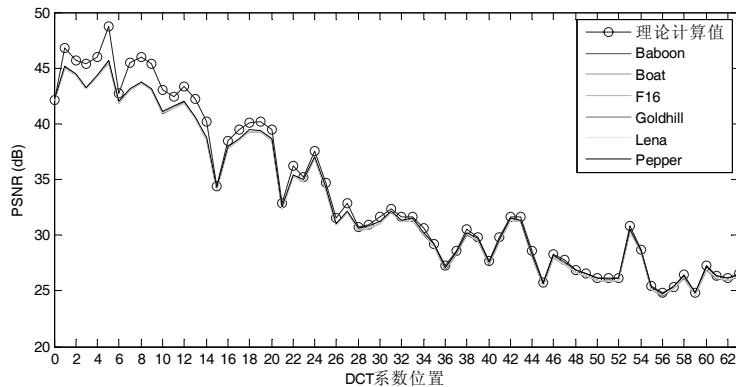


图 5.22 按原始 JPEG 图像和伪装 JPEG 图像间 PSNR 评估每个量化 DCT 系数对图像质量的影响

通过比较图 5.22 与图 5.23 可以看出，两个图走势大致相同。Baboon 图由于有较复杂的纹理，PSNR 值低于其余图像。从这两个图可以看出，由于量化步骤的引入，低频量化 DCT 系数甚至 DC 系数对图像质量的影响比中高频系数要来得低。由于量化 DC 系数是按相邻分块块差值编码的，修改 DC 系数对文件大小的变化影响不大；而量化 AC 系数是按 ZRLE (Zlib Run-Length Encoding) 进行编码，中高频 AC 的改变会导致连零的中断从而可能使码流变长。因此，当需要嵌入较多信息时，优先按 zigzag 顺序选择量化

DC 系数和低频 AC 系数将有利于达到较高的图像质量和较少的文件大小增长。不过，DC 量化系数的改变虽然以 PSNR 评估并不会造成多大的图像质量下降，但实际视觉效果会容易导致块效应的出现。研究表明，量化 AC 系数出现连续多个 0 的情况，对于质量因子 QF 不大于 85 的情况，一般从第 6 个量化 AC 系数附近开始，而其余情况一般从第 17 个左右开始。综合以上信息，选择第 3 到第 14 个量化 AC 系数来嵌入信息。

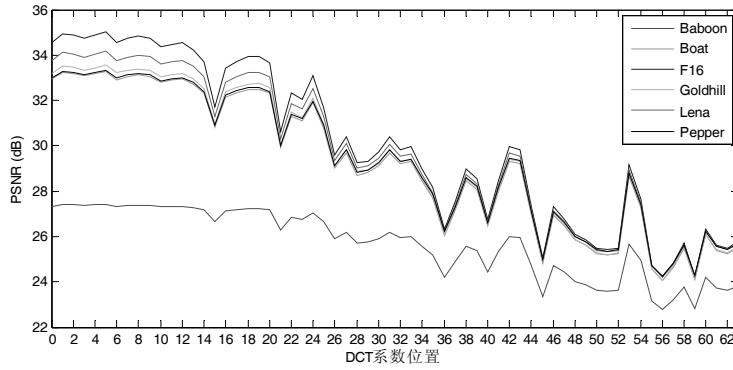


图 5.23 按原始未压缩图像和伪装 JPEG 图像间 PSNR 评估每个量化 DCT 系数对图像质量的影响

(2) 嵌入步骤

由文献[84]所提出的方法，可以联想到将其二进制推广到 k 进制。为了在量化 DCT 系数中产生冗余空间来嵌入信息，我们对某些量化表元素进行整除，同时提升对应的量化 DCT 系数。当一个整数 M 乘以一个整数 k ，结果是 k 的倍数，这就使得我们可将按 k 进制表示的秘密信息或附加信息嵌入其中，采用模操作 $M \bmod k$ 即可提取出嵌入的信息。

假设在 (u, v) 位置的原始量化表元素为 $Q(u, v)$ ，对其除以 k 以获得一个新的量化表元素。由于量化表元素是整数，为了使量化 DCT 系数上的改变在图像质量上的影响尽可能小，对整除选择向下取整

$$Q'(u, v) = \lfloor Q(u, v) / k \rfloor \quad (5.46)$$

其中 $\lfloor \cdot \rfloor$ 表示向下取整。然后对应的量化 DCT 系数 $D(u, v)$ 乘以 k ，即 $D'(u, v) = k \times D(u, v)$ ，至此生成了额外的嵌入空间。由于向下取整函数可能会导致信息的丢失，因此在 $D'(u, v)$ 上增加一个调整量 X 来使 $D'(u, v) \times Q'(u, v)$ 尽可能地接近原始的重构 DCT 系数 $\tilde{F}(u, v)$ 。假设 $r(u, v)$ 是 $Q(u, v)/k$ 的余数，则有 $r(u, v) = Q(u, v) - k \times Q'(u, v)$ 。为了尽可能地达到 $\tilde{F}(u, v) = D(u, v) \times Q(u, v) = D'(u, v) \times Q'(u, v)$ 的目的，可令

$$D(u, v) \cdot [k \cdot Q'(u, v) + r(u, v)] = [k \cdot D(u, v) + X] \cdot Q'(u, v) \quad (5.47)$$

得到

$$X = \frac{r(u, v) \cdot D(u, v)}{Q'(u, v)} \quad (5.48)$$

X 可能不是 k 的倍数，因此令

$$X = k \cdot \text{round}\left(\frac{r(u, v) \cdot D(u, v)}{k \cdot Q'(u, v)}\right) \quad (5.49)$$

从而得到

$$D'(u, v) = k \cdot D(u, v) + k \cdot \text{round}\left(\frac{r(u, v) \cdot D(u, v)}{k \cdot Q'(u, v)}\right) \quad (5.50)$$

基于上述思想, 详细嵌入步骤可以描述如下。

步骤 1: 对原始 JPEG 图像部分解码得到量化表和量化 DCT 系数, 同时得到图像大小以计算 8×8 分块的个数 N 。

步骤 2: 加密原始秘密信息, 并转换成 k 进制, 计算转换后的 k 进制数字个数 N_D 。这一步可使得嵌入容量大大增加。

步骤 3: 计算需要应用式 (5.46) 进行修改的量化表元素个数 N_Q 。一个量化 DCT 系数能嵌入一个 k 进制数字信息, 一个量化表元素对应 N 个量化 DCT 系数。可按下式计算 N_Q

$$N_Q = \lceil (N_D + 24) / N \rceil \quad (5.51)$$

其中 $\lceil \cdot \rceil$ 表示向上取整, 24 是额外信息位数。接着以 zigzag 顺序按式 (5.46) 修改量化表, 从第三个元素开始修改直到改完 N_Q 个元素。为了后续的无损恢复, 应用式 (5.46) 后的余数 $r(u, v)$ 需要作为额外信息记录下来, 共需 12 位; 而有时结果为零, 则跳过该元素, 保持其原始数值, 同时需要一个标志位作为额外信息表示该元素是否修改, 也需 12 位。所以式 (5.51) 中需要加 24。还有, 第一个修改的量化表元素位置也要保存下来, 亦可作为提取密钥。

步骤 4: 置乱 DCT 块 (8×8 DCT 块作为一个整体, 内部不置乱), 然后将秘密信息连同额外信息一起嵌入。信息嵌入置乱后的分块, 当分块还原为原来的位置时, 嵌入信息带来的改变能够较为均匀地分布在整个图像。打散机制让修改的分块分散, 使其不会集中于图像的某一部分, 因为图像质量的退化集中于某个局部会使图像看起来不协调。对于嵌入方式, 首先将一个数字位嵌入第 i 个分块的第 j 个量化 DCT 系数, 然后将下一个数字位嵌入第 $i+1$ 个分块的第 j 个量化 DCT 系数; 当所有分块的第 j 个量化 DCT 系数都已嵌入, 再开始往每个分块的第 $j+1$ 个量化 DCT 系数嵌入。图 5.24 给出了一个 16×16 大小的 JPEG 图像嵌入方式示例, 从置乱后的第一个 DCT 分块第三个量化 DCT 系数开始嵌入。假设转换后的 k 进制位为 W , 为了让图像质量下降尽可能低, 以使改变尽量在以 $D'(u, v)$ 为中心的一个范围内, 采用以下嵌入规则

$$D''(u, v) = \begin{cases} D'(u, v) + W & W \leq k/2 \\ D'(u, v) + W - k & \text{其他} \end{cases} \quad (5.52)$$

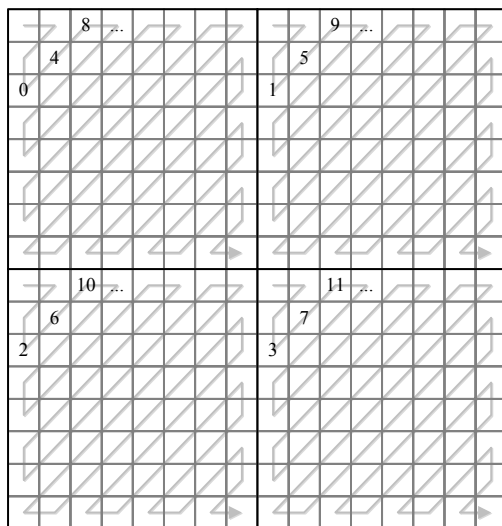


图 5.24 16×16 大小的 JPEG 图像嵌入方式示例

信息嵌入后, 所有分块再还原到原来的位置。

步骤 5: 将修改的量化表保存到 JPEG 的文件头, 重新对修改的量化 DCT 块进行熵编码, 得到伪装 JPEG 图像。

(3) 提取和恢复步骤

信息提取和图像恢复过程是信息嵌入过程的逆过程, 详细提取和恢复步骤如下。

步骤 1: 解码伪装 JPEG 图像得到量化表和量化 DCT 分块。

步骤 2: 按嵌入步骤相同的置乱方式置乱 DCT 分块, 根据提取密钥得知第一个伪装量化 DCT 系数位置。采用模操作提取嵌入的信息:

$$W = D''(u, v) \bmod k \quad (5.53)$$

标志位和余数 $r(u, v)$ 等额外信息首先提取出来, 紧跟着是秘密信息。同时, 量化 DCT 系数采用下式进行恢复

$$D'(u, v) = \begin{cases} D''(u, v) - W & W \leq k/2 \\ D''(u, v) - W + k & \text{其他} \end{cases} \quad (5.54)$$

$$D(u, v) = \text{round}\left(\frac{D'(u, v)}{k + r(u, v) / Q'(u, v)}\right) \quad (5.55)$$

式 (5.55) 的证明如下: 由公式 (5.50) 和四舍五入的定义可知

$$D'(u, v) = k \cdot D(u, v) + \frac{r(u, v) \cdot D(u, v)}{Q'(u, v)} + k \cdot R \quad (5.56)$$

其中 $R \in (-\frac{1}{2}, \frac{1}{2}]$ 。也就是说

$$D(u, v) = \frac{D'(u, v) - k \cdot R}{k + r(u, v) / Q'(u, v)} \quad (5.57)$$

其等价于

$$D(u, v) + \frac{k \cdot R}{k + r(u, v) / Q'(u, v)} = \frac{D'(u, v)}{k + r(u, v) / Q'(u, v)} \quad (5.58)$$

因为 $\frac{k \cdot R}{k + r(u, v) / Q'(u, v)} \in \left(-\frac{1}{2}, \frac{1}{2}\right)$, 即

$$D(u, v) - \frac{1}{2} < \frac{D'(u, v)}{k + r(u, v) / Q'(u, v)} < D(u, v) + \frac{1}{2} \quad (5.59)$$

根据四舍五入定义有 $\text{round}(x) = n \Leftrightarrow n - \frac{1}{2} \leq x < n + \frac{1}{2}$, 由此式 (5.55) 证毕。

步骤 3: 将步骤 2 提取的秘密信息转换回二进制, 并进行解密得到原始秘密数据。至此, 提取部分完成。

步骤 4: 利用额外信息 $r(u, v)$ 恢复原始量化表

$$Q(u, v) = k \cdot Q'(u, v) + r(u, v) \quad (5.60)$$

步骤 5: 将恢复的原始量化表保存在 JPEG 文件头中, 对恢复的原始量化 DCT 块进行熵编码, 得到原始 JPEG 图像。

3. 实验结果与分析

为了讨论文献[91]方案的性能, 我们分析了文献[91]方案的特性并同其他方案进行比较。这里采用大小为 512×512 的 256 灰度 Lena 图像进行测试, 从图像质量、嵌入容量和文件大小三个方面进行讨论, 其中图像质量采用 PSNR 进行评估。

(1) 算法性能

秘密信息的 k 进制转换使得嵌入容量大大增加, 但不同 k 值情况下性能可能不太一样, 我们采用了几个不同的 k 值进行实验。对于 $k = 3$, 我们将三个二进制位转为两个三进制位, 相当于 1 个三进制位可以嵌入 1.5 个二进制位信息; 对于 $k = 4$, 1 个四进制位可以嵌入 2 个二进制位信息, 对于其他 k 情况类似。

表 5.3 给出了不同 k 值下的嵌入容量、图像质量和文件大小的性能情况, 实验对象是 QF=70 的 Lena 图像。观察嵌入容量可以看出, 随着 k 值的增大嵌入容量刚开始跟着增大, 但当 k 值大于 6 时, 嵌入容量开始波动。某些量化表元素其值较小, 当 k 值较大时不足以应用式 (5.46) 而保持原值, 对应量化 DCT 系数也就保持不变以至于没有产生可用于嵌入信息的冗余空间。因此, k 值的选择不能太大, 文献[91]选择 2、3 和 4 作为 k 值继续进行实验。

表 5.3 QF 为 70 的 Lena 图像在不同 k 值下的性能

k	嵌入容量 (bit)	PSNR (dB)		文件大小 (KB)	
		0.125 bpp	全嵌入	0.125 bpp	全嵌入
2	49128	35.61	35.39	38.2	40.4
3	73692	35.80	35.64	39.0	45.0
4	98256	35.80	35.58	40.9	49.6
5	112000	35.82	35.60	39.8	52.9
6	126252	35.83	35.63	40.4	55.1
7	113477	35.84	35.60	38.5	53.5
8	122808	35.82	35.63	39.0	52.8

表 5.4、表 5.5 和表 5.6 给出了不同嵌入率和不同质量因子条件下 Lena 图像的实验结果, 左斜杠 (/) 表示不能达到该嵌入率。从这三个表中可以看出, 当嵌入率越大或质量因子 QF 越小, PSNR 越低。比较这几个表, 修改相同位数的 PSNR 是类似的; 但由于是 k 进制, 嵌入率却是成倍的。为了更好地比较不同 k 值对图像质量的影响, 图 5.25 给出了不同嵌入率下 QF 为 70 的 Lena 图像质量对比图, 在嵌入率相同的情况下, $k = 3$ 和 $k = 4$ 的结果比 $k = 2$ 的结果要好。

表 5.4 不同嵌入率和不同品质因数下 Lena 图像的图像质量 ($k = 2$)

QF	嵌入率 ($\times 1$ bpp)					
	0.031	0.062	0.094	0.125	0.156	0.187
50	34.32	34.20	34.07	33.90	33.74	33.48
60	34.97	34.89	34.80	34.69	34.55	34.37
70	35.84	35.76	35.70	35.63	35.53	35.39
80	37.00	36.97	36.93	36.89	36.83	36.73
90	39.43	39.41	39.39	39.36	39.34	39.29

表 5.7 是不同 k 值下的嵌入容量, 可以很明显地看出 k 值较大时嵌入容量也较大, 但当品质因数 QF 变得很大时造成大量的量化表元素不足以应用式 (5.46) 而保持原值, 因此对应量化 DCT 系数也保持不变, 没有产生可用于嵌入信息的冗余空间。

图 5.26 是 QF 为 70 的 Lena 图像嵌入信息前后文件大小的比较图, 随着嵌入率的增加文件大小的整体趋于线性增加。一般来说, 我们嵌入越多数据, 需要修改越多的量化

表元素和量化 DCT 系数, 而 JPEG 图像是因 AC 系数采用 ZRLE 压缩而使其文件大小大大降低的。文献[91]选择第 3 到第 14 个量化 DCT 系数进去数据的嵌入, 有效地防止了中高频连零 AC 系数被破坏, 在一定的嵌入率下文件大小的增长是可接受的。

从上述分析可以看出, 在同样条件下参数 k 值越大, 图像质量越好, 也能够达到更大的嵌入容量, 不过, k 值的增大也将导致文件大小的增长。因此, k 值的选择需要在图像质量和文件大小间找到一个平衡点。需要注意的是, 当载体 JPEG 图像品质因数 QF 较高时 k 值不能太大, 否则嵌入容量将会不升反降。

(2) 同现有方法比较

为了更好地展现文献[91]方法相对于现有方法的优越性, 我们将文献[91]方法同 Chang 等人的方法^[88]及 Xuan 等人的方法^[89]进行对比, 主要在图像质量和文件大小两方面进行比较。为保证公平性, 实验统一采用品质因数为 70 的 Lena 图像, 文献[91]方法参数 k 值为 2。

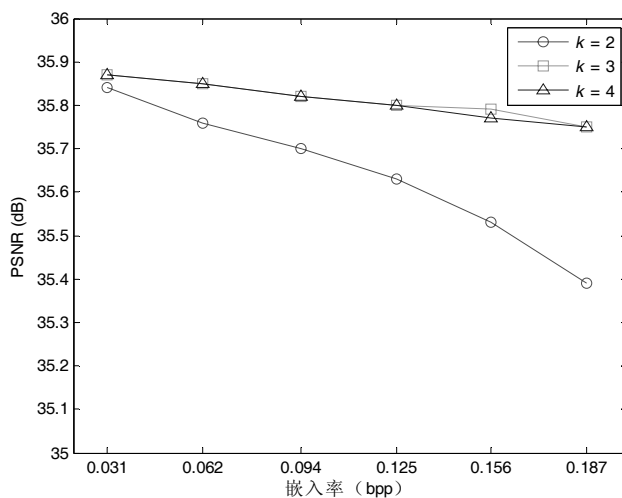


图 5.25 不同 k 值在 QF 为 70 条件下的 Lena 图像质量对比

表 5.5 不同嵌入率和不同品质因数下 Lena 图像的图像质量 ($k=3$)

QF	嵌入率 ($\times 1.5$ bpp)					
	0.031	0.062	0.094	0.125	0.156	0.187
50	34.38	34.31	34.24	34.14	34.06	33.89
60	35.01	34.96	34.92	34.87	34.80	34.68
70	35.86	35.82	35.80	35.76	35.71	35.64
80	37.02	37.00	36.98	36.96	36.92	36.88
90	39.43	39.42	39.40	39.39	39.36	/

表 5.6 不同嵌入率和不同品质因数下 Lena 图像的图像质量 ($k=4$)

QF	嵌入率 ($\times 2$ bpp)					
	0.031	0.062	0.094	0.125	0.156	0.187
50	34.36	34.26	34.19	34.09	33.97	33.77
60	35.00	34.94	34.90	34.83	34.75	34.62
70	35.85	35.80	35.75	35.71	35.66	35.58
80	37.00	36.98	36.94	36.90	36.86	36.79
90	39.41	/	/	/	/	/

表 5.7 不同 k 值和品质因数下 Lena 图像的嵌入容量

QF	70	75	80	85	90
$k=2$	49128	49128	49128	49128	49128
$k=3$	73692	73692	73692	73692	61404
$k=4$	98256	98256	98256	90064	24528

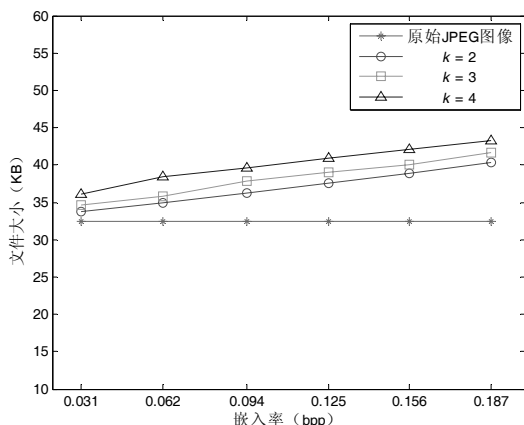


图 5.26 QF 为 70 的 Lena 图像嵌入信息前后文件大小比较

图 5.27 是不同嵌入率下的图像质量比较图，从图中可以看出文献[91]方法（图中“本文方法”）的 PSNR 值要大大好于其他两种方法，有更好的图像质量。一个原因是文献[91]方法对原始 DCT 系数的改变不大于 $\frac{1}{2}Q(u,v)$ ，而 Chang 等人的方法和 Xuan 等人的方法都是 $Q(u, v)$ 。Chang 等人的方法 PSNR 较低是因为其将信息嵌入在 DCT 分块的中频零系数中，根据前面的分析，低频量化 DCT 系数的改变对图像质量的影响小于中频和低频系数。Xuan 等人的方法同样选择中低频系数进行信息嵌入，不过其采用直方图对扩展技术导致需要更多系数。当需要嵌入较多信息时，需要更多直方图对进行扩展导致图像质量快速下降。

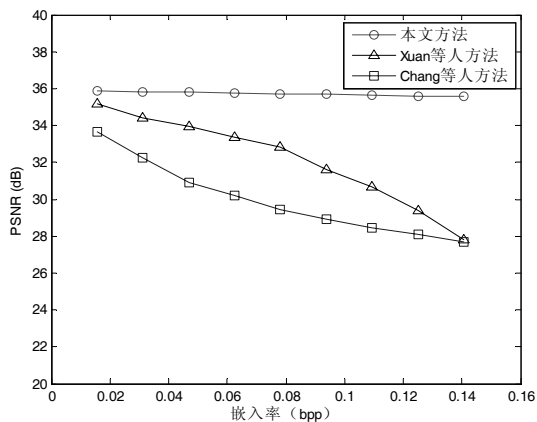


图 5.27 文献[91]方法同现有方法图像质量比较

图 5.28 是文件大小的比较图，嵌入率为 0.125 bpp。对于不同品质因数，文献[91]方法（图中标“本文方法”）及 Xuan 等人方法得到的伪装图像的文件大小更接近原始 JPEG 文件，而 Chang 等人的方法则较大于原始文件大小，主要也是因为 Chang 等人的

方法将信息嵌入在 DCT 分块的中频零系数中,破坏了连零从而影响到 ZRLE 的压缩效率。当品质因数较大时,三者的文件大小趋于原始 JPEG 文件大小,因为高品质因数意味着较低的压缩率,导致上述结果。

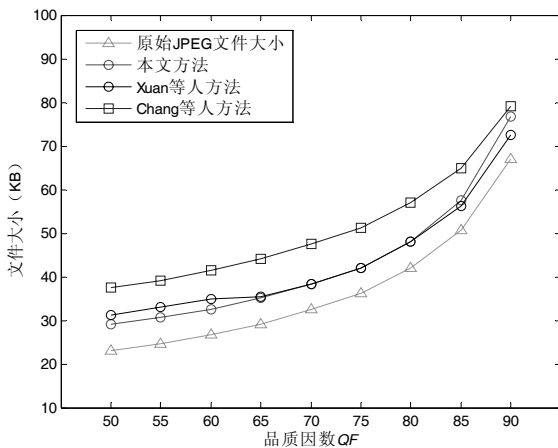


图 5.28 文献[91]方法同现有方法文件大小比较

5.7 本章小结

无损信息隐藏技术是隐写术和数字水印的一个重要分支。无损信息隐藏可以无失真的提取嵌入信息以及完整恢复原始载体,它可以应用在版权信息保护,军事、医用图像以及图像取证等诸多方面,因而无损信息隐藏得到了研究人员的极大关注。本章首先介绍无损信息隐藏技术的提出背景、相关概念、关键问题和分类,接着介绍无损信息隐藏的框架模型和性能评价问题。然后以图像载体为背景,分空域、变换域和压缩域三大类介绍一些典型的无损信息隐藏技术。对于空域,主要介绍了四种典型方法:基于无损压缩替换的、基于模加的、基于差值扩展的以及基于直方图移位的。对于变换域,主要介绍基于整数 DCT 变换域和整数 DWT 变换域的无损信息隐藏技术。同空域直方图移位算法相比,整数变换域直方图移位算法的性能优势显著。基于变换域的数值扩展和直方图移位算法要比空域算法性能更稳定,空域算法受像素分布影响较大,对不同统计特征的图像性能不够稳定。对于压缩域,主要介绍基于 BTC 和 JPEG 压缩域的无损信息隐藏技术。当从压缩域的角度考虑无损信息隐藏算法时,所关心的对象实际上已从原始图像的像素值形式转为压缩域的码流形式。该嵌入对象的转变也使无损信息隐藏嵌入算法所对应的技术和应用要求有所转变。



习题

1. 请阐述无损信息隐藏的技术难点和性能评价指标。
2. 分别以 $\{x=201, y=206, b=1\}$ 和 $\{x=200, y=246, b=1\}$ 为例,请解释学者 Tian 提出的差值扩展方法,并指出相应的差值是否可扩展,是否可交换。
3. 请用 Matlab 或 C 语言编写一个程序,实现直方图移位算法,在 256 灰度 Lena 图像中嵌入 100 比特随机信息,然后提取出信息并恢复原始图像。

4. 学者 Alattar 提出一种可逆整数变换函数 (Reversible Integer Transform Function), 用来无损信息隐藏。他把图像分割成一系列块, 每个块可以看成是一个矢量, 然后利用可逆整数变换函数对矢量进行变换, 如果变换得到的矢量是可以用来隐藏数据的, 则利用 Tian 的方法进行隐藏。举一个例子说明, 假设有一个向量 $c=(c_1, c_2, c_3, c_4)$ 里面包含有四个像素, 我们在该矢量内隐藏三比特 m_1, m_2, m_3 。首先, Alattar 使用可逆整数变换函数计算矢量的加权平均 v_1 , 得到新的矢量 $v=(v_1, v_2, v_3, v_4)$, 方程式如下所示

$$v_1 = \left\lfloor \frac{a_1 c_1 + a_2 c_2 + a_3 c_3 + a_4 c_4}{a_1 + a_2 + a_3 + a_4} \right\rfloor, v_2 = c_2 - c_1, v_3 = c_3 - c_1, v_4 = c_4 - c_1 \quad (5.61)$$

其中 a_1, a_2, a_3, a_4 是四个常数值。接着, 依据秘密信息 m_1, m_2, m_3 来调整 v , 如下所示

$$v'_1 = v_1, v'_2 = 2 \times v_2 + m_1, v'_3 = 2 \times v_3 + m_2, v'_4 = 2 \times v_4 + m_3 \quad (5.62)$$

最后, 产生伪装像素 $s=(s_1, s_2, s_3, s_4)$

$$s_1 = v'_1 - \left\lfloor \frac{a_2 v'_2 + a_3 v'_3 + a_4 v'_4}{a_1 + a_2 + a_3 + a_4} \right\rfloor, s_2 = v'_2 + s_1, s_3 = v'_3 + s_1, s_4 = v'_4 + s_1 \quad (5.63)$$

在提取秘密信息和恢复像素的过程中, 先根据伪装矢量 s , 计算矢量 $u=(u_1, u_2, u_3, u_4)$ 如下所示

$$u_1 = \left\lfloor \frac{a_1 s_1 + a_2 s_2 + a_3 s_3 + a_4 s_4}{a_1 + a_2 + a_3 + a_4} \right\rfloor, u_2 = s_2 - s_1, u_3 = s_3 - s_1, u_4 = s_4 - s_1 \quad (5.64)$$

接着由矢量 u 推导出隐藏的信息

$$m_1 = u_2 - \left\lfloor \frac{u_2}{2} \right\rfloor, m_2 = u_3 - \left\lfloor \frac{u_3}{2} \right\rfloor, m_3 = u_4 - \left\lfloor \frac{u_4}{2} \right\rfloor \quad (5.65)$$

接着, 计算矢量 u'

$$u'_1 = u_1, u'_2 = \left\lfloor \frac{u_2}{2} \right\rfloor, u'_3 = \left\lfloor \frac{u_3}{2} \right\rfloor, u'_4 = \left\lfloor \frac{u_4}{2} \right\rfloor \quad (5.66)$$

最后, 原始的像素值利用下面的式子还原

$$c_1 = u'_1 - \left\lfloor \frac{a_2 u'_2 + a_3 u'_3 + a_4 u'_4}{a_1 + a_2 + a_3 + a_4} \right\rfloor, c_2 = u'_2 + u'_1, c_3 = u'_3 + u'_1, c_4 = u'_4 + u'_1 \quad (5.67)$$

请用 $(c_1, c_2, c_3, c_4)=(8, 12, 15, 12)$, $(m_1, m_2, m_3)=(1, 0, 1)$, $(a_1, a_2, a_3, a_4)=(1, 2, 2, 1)$ 来验证上述方法的可行性。

5. 请用 Matlab 或 C 语言编写一个程序, 实现 Chen 等人提出的 AMBTC 压缩域无损信息隐藏算法, 分别以 256 灰度 Lena 图像和 Baboon 图像做实验, 比较这两幅图像的嵌入容量。

其他信息隐藏研究分支简介

本章引言

前面第 2、3、4 章已经介绍了信息隐藏技术的三大主要分支：隐写术、数字水印和数字指纹。第 5 章提到的无损信息隐藏则是隐写术和数字水印中的特殊情况。除此之外，还存在其他一些研究分支，包括隐蔽信道、阈下信道、低截获概率通信和匿名通信，本章分别对它们进行简要介绍。

本章重点

- 隐蔽信道的概念和研究分支；
- 阈下信道的概念、模型和构造方法；
- 扩频通信的方式；
- 匿名通信系统体系结构。



6.1 隐蔽信道

隐蔽信道 (Covert Channel) 是指允许进程以危害系统安全策略的方式传输信息的通信信道, 是对安全信息系统的重要威胁。隐蔽信道普遍存在于安全操作系统、安全网络、安全数据库系统中。我国的《计算机信息系统安全保护等级划分准则》(GB17859.1999)、美国的《可信计算机系统评估准则》(TCSEC) 及国际标准化组织 ISO 在 1999 年发布的《信息技术安全评估通用准则》(ISO / IEC15408, 简称 CC 标准) 都要求对高等级安全信息系统进行隐蔽信道分析, 并在识别隐蔽信道基础上, 对隐蔽信道进行度量和处置。本节先介绍隐蔽信道的基本概念、分类、研究领域, 然后简要介绍隐蔽信道分析技术^[92]。

6.1.1 隐蔽信道基本概念

隐蔽信道的概念最初由 Lampson 于 1973 年提出, 他给出的隐蔽信道定义为: 不是特意设计或本意不是用来传输信息的通信信道。当时 Lampson 主要关注程序的限制问题, 即如何在程序的执行过程中进行限制, 使其不能向其他未授权的程序传输信息。他列举了恶意或行为不当的程序绕过限制措施、泄露数据的六种方法和相应的处理措施, 并把这些方法归纳为 3 种类型: 存储信道、合法信道和隐蔽信道。后续的研究将隐蔽信道重新划分为两种类型: 存储隐蔽信道和时间隐蔽信道, 统称隐蔽信道。时间隐蔽信道对应于 Lampson 所指的隐蔽信道; 合法信道则是一种阈下信道, 是公开信道中所建立的一种实现隐蔽通信的方式, 信道中公开的、有意义的信息仅仅充当了秘密信息的载体, 秘密信息通过它进行传输。这种隐蔽传输信息的方式后来逐渐淡出了隐蔽信道研究的中心, 形成了相对独立的研究领域, 因此将在下一小节单独介绍。与利用隐蔽信道进行信息隐藏相对立的就是隐蔽信道分析工作, 包括信道识别、度量和处置。信道识别是对系统的静态分析, 强调对设计和代码进行分析发现所有潜在的隐蔽信道。信道度量是对信道传输能力和威胁程度的评价。信道处置措施包括信道消除、限制和审计。隐蔽信道消除措施包括修改系统、排除产生隐蔽信道的源头、破坏信道的存在条件。限制措施要求将信道危害降低到系统能够容忍的范围内。但是, 并非所有的潜在隐蔽信道都能被入侵者实际利用, 如果对所有的潜在隐蔽信道进行度量和处置会产生不必要的性能消耗, 降低系统效率。隐蔽信道检测则强调对潜在隐蔽信道的相关操作进行监测和记录, 通过分析记录, 检测出入侵者对信道的实际使用操作, 为信道度量和处置提供依据。

在隐蔽信道研究过程中, 学者们给出了多种不同的定义。比较全面的隐蔽信道定义如下: 给定一个强制安全策略模型 M 及其在一个操作系统中的解释 $I(M)$ 。 $I(M)$ 中的两个主体 $I(S_h)$ 和 $I(S_l)$ 之间的通信是隐蔽的, 当且仅当模型 M 中的对应主体 S_h 和 S_l 之间的任何通信都是非法的。该定义指出, 隐蔽信道只与系统的强制访问控制策略模型相关。隐蔽信道广泛存在于部署了强制访问控制机制的安全操作系统、安全网络和安全数据库中。

1973 年, Bell 和 LaPadula 提出了著名的 Bell-LaPadula 多级安全强制访问控制模型 (BLP 模型), 该模型描述如下: 系统包含主体集 S 和客体集 O , S 中的每一个主体 s 和 O 中的每一个客体 o 都分别具有一个固定的安全标记 $C(s)$ 和 $C(o)$ (表示信任和敏感等级)。BLP 模型在安全标记之间建立了一种称为“支配”的偏序格关系, 用“ \geq ”表示。BLP 模型要求安全系统内主体操作客体的安全信息流具有简单安全特性和*-特性: ① 简单安全特性: 仅当 $C(s) \geq C(o)$ 时, 主体 s 才可以对客体 o 有“读”访问权限; ② *-特性

(Star Property): 仅当 $C(p) \geq C(o)$ 时, 对客体 o 有“读”访问权限的主体才可以对客体 p 有“写”访问权限。简单安全特性表明, 信息接收者的信任等级不得低于信息的敏感等级。*-特性表明, 将一个敏感对象的内容写入另一个敏感对象, 要求后者的敏感等级至少不低于前者。这两个特性可简单概括为“不向上读, 不向下写”或叫做“上面读, 下面写”。BLP 模型既可以阻止低级别的主体访问高密级的信息, 同时也阻止高安全级别主体通过“写”操作向低级别主体泄漏信息。

即使在强制访问控制模型下, 恶意用户仍然能够通过构建隐蔽信道实现从高安全级主体向低安全级主体的信息传输, 实现方式如图 6.1 所示。高安全级和低安全级用户之间通过修改和感知共享变量的值或者属性传递信息。TCSEC 标准使用 TCB (Trusted Computing Base, 可信计算基) 表示计算机系统中所有保护机制的总和 (包括硬件、固件和软件), 负责执行安全策略。因此, 隐蔽信道可以表示为 TCB 三元组

$$\langle \text{variable}, \text{PA}_h, \text{PV}_i \rangle \quad (6.1)$$

其中, variable 是系统中的变量; PA_h 是修改这个变量的 TCB 原语且具有较高的安全级; PV_i 是感知、观察这个变量的 TCB 原语且安全级较低。从 PA_h 到 PV_i 的通信是系统安全策略所不允许的, 则 PA_h 到 PV_i 的通信信道称为隐蔽信道。

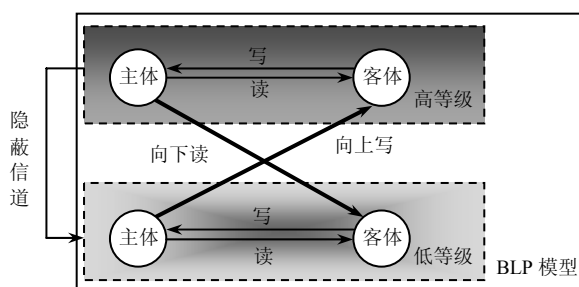


图 6.1 隐蔽信道示例

6.1.2 隐蔽信道分类

隐蔽信道三元组 $\langle \text{variable}, \text{PA}_h, \text{PV}_i \rangle$ 中变量 variable 可以表示系统中不同的属性, 当 variable 表示存储属性时, 隐蔽信道为存储隐蔽信道。例如, 在资源耗尽型信道中, variable 表示收发双方能够修改和感知的共享资源。当 variable 表示 CPU 时间或者响应时间等属性时, 隐蔽信道为时间隐蔽信道。在对隐蔽信道识别方法的研究中, 学者们给出了各种隐蔽信道存在的最小条件。

存储隐蔽信道存在的最小条件如下。

- (1) 信息的发送者和接收者必须能够访问某个共享资源的同一个属性;
- (2) 信息的发送者能够以某种方式改变这个属性;
- (3) 同时, 信息的接收者必须能够检测这个属性的任何一个改变;
- (4) 存在着某种机制初始化发送者和接收者, 并且要保证发送和接收时间顺序的正确性, 即建立好的同步机制以保证信息正确地发送与接收。

时间隐蔽信道存在的最小条件如下。

- (1) 发送者和接收者必须对某个共享资源的同一个属性有访问权;
- (2) 发送者和接收者必须有一个统一的时间参考, 比如一个实际时钟;
- (3) 发送者必须能够调制接收者的响应时间来表示一个属性的改变;

(4) 一定存在某个机制使得发送和接收双方能够同步发送事件。

与存储隐蔽信道相比, 时间隐蔽信道又称为无记忆通道, 不能长久地存储信息。发送者发送的信息接收者必须及时接收, 否则要传递的信息就会消失, 时效性较强。分析隐蔽信道存在条件及其表示 $\langle \text{variable}, \text{PA}_h, \text{PV}_i \rangle$ 可知, 存储隐蔽信道和时间隐蔽信道并没有本质的区别, 只是 variable 变量代表的属性不同。在隐蔽信道分析中, 时间隐蔽信道具有更大的复杂性, TCSEC 标准要求 B2 级安全系统进行彻底的存储隐蔽信道分析, 而更高级别的 B3 级和 A1 级安全系统才要求必须同时进行时间隐蔽信道分析。

与其他的通信信道类似, 隐蔽信道也可以分为噪音信道和无噪信道。对于 $\langle \text{variable}, \text{PA}_h, \text{PV}_i \rangle$ 中的 variable 变量, 如果该变量只能被 PA_h 原语修改, 而且对于任意修改, PV_i 原语都能够实现概率为 1 的正确解码, 则该信道称为无噪信道; 如果 variable 变量被 PA_h 原语修改的同时还可能被其他原语修改, PV_i 不能正确解码, 则该信道称为噪音信道。在隐蔽信道分析中, 通常将信道抽象成无噪信道以度量信道最大容量。但是, 在实际场景中, 信道多为噪音信道, 影响隐蔽信道的传输效率。

隐蔽信道传输有固定的信息传输周期, 如图 6.2 所示。一个完整的信息传输周期包括发送者 / 接收者同步阶段、信息传输阶段和反馈阶段。在同步阶段中, 发送者通知接收者准备发送信息, 如果收发双方有事先的约定, 例如每隔 t 个时间单元发送新的信息, 则同步阶段可以省略; 如果收发双方通信路径不可信, 则反馈阶段必须存在, 否则发送者无法确认接收方是否收到信息, 也无法确认何时开启新的传输周期。

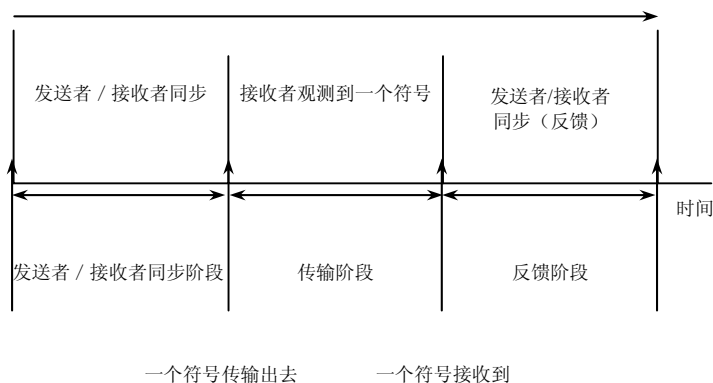


图 6.2 隐蔽信道周期

6.1.3 隐蔽信道研究领域

Lampson 提出的隐蔽信道概念关注于程序的限制问题, 但当前的隐蔽信道问题已经涉及安全信息产品的多个领域, 包括安全操作系统、安全数据库、安全网络等。此外, 还存在一些倍受关注的与隐蔽信道相关联的概念。下面分析隐蔽信道涉及到的研究领域以及相互之间的区别和联系, 进一步加深对隐蔽信道的理解。

1. 数据库隐蔽信道

数据库系统中的隐蔽信道主要包括以下三类:

(1) 数据库存储资源引入的信道

信道可利用的数据库存储资源包括数据和数据字典, 其主要原理是发送者修改数据/数据字典, 接收者则通过完整性约束等方式间接感知数据/数据字典的修改, 从而获得信息;

(2) 数据库管理资源引入的信道

主要是数据库系统变量或安全机制的资源耗尽型信道。资源包括游标、临时数据区等。另外，同时管理多个安全级别用户的系统安全机制也可能引入信道，如审计机制等。

(3) 事务并发控制引起的隐蔽信道

安全数据库系统中通常依据 BLP 模型实施强制访问控制，约束用户的数据访问操作，以保证数据的安全性。同时，为了保证数据操作的实时性，系统还需要采用实时算法处理事务调度和并发控制。恶意主体可以利用不同安全级事务间的并发冲突构造隐蔽信道，称作**数据冲突隐蔽信道**（Data Conflict Covert Channel，简称 DC 信道）。

在数据冲突隐蔽信道中，两个不同安全级别的用户发起的事务 tr_l 、 tr_h 共同访问同一数据项 d_x ，其安全级别关系为 $SL(tr_h) \geq SL(d_x) \geq SL(tr_l)$ 。其中，低安全级别事务 tr_l 写访问 d_x ，高安全级别事务 tr_h 读访问 d_x 。在该场景下，可以构造多种具体的数据冲突隐蔽信道，实现高安全级用户向低级别用户传递信息。例如，入侵者利用事务执行过程中是否发生冲突的事实表示希望传输的符号，如图 6.3 所示。首先，低安全级用户发起事务 tr_l ，如果高安全级用户希望发出符号‘1’，则发起事务 tr_h 。由于两个事务间存在冲突，系统放弃事务 tr_l ；如果高安全级用户希望发出符号‘0’，则不发起事务 tr_h ， tr_l 可以执行完成。在该场景下，处于不同安全级别的事务通过并发控制机制相互干扰传递信息。

2. 阈下信道

阈下信道（Subliminal Channel）是指在基于公钥密码技术的数字签名、认证等应用密码体制的输出密码数据中建立起来的一种隐蔽信道，除指定接收者外，任何其他人均不知道密码数据中是否存在阈下消息。阈下信道又称潜信道，是一种信息隐藏方法。Simmons 于 1984 通过研究看守监狱中两个囚犯秘密协商逃跑计划的例子（图 6.4），引入阈下信道的概念。图 6.4 中，Wendy 负责监视囚犯 Alice 和 Bob 的活动，一旦发现异常就将其分开并更加严格看管；Alice 和 Bob 必须采用某种约定协商越狱情况。

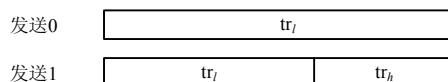


图 6.3 数据冲突隐蔽信道

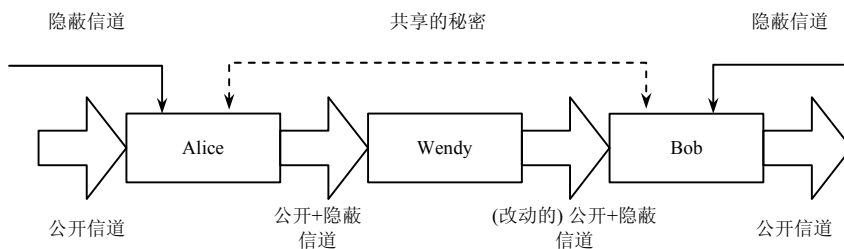


图 6.4 囚犯问题

虽然在 Simmons 的定义中阈下信道也被称作隐蔽信道，但是一般认为，阈下信道与 TCSEC 标准中所指的隐蔽信道有所差别，具体如下。

(1) 阈下信道是在公开信道中所建立的一种实现隐蔽通信的信道。由于传输信道本身的合法性，其更接近于信息隐藏研究范畴，且已经被认可为一种典型的信息隐藏方法；

(2) 阈下信道的宿主是密码系统，且只能是非对称密码体制。阈下信道的特点是，即使监视者知道要寻求的内容，也无法发现信道的使用并获取正在传送的阈下消息。因

为阙下信道的加密特性决定了其安全性要么是无条件的，要么是计算上不可破的。这不是普通隐蔽信道所能做到的而且也不是必需的。

有关阙下信道的详细介绍见 6.2 节。

3. 网络信道

网络信道一般可以分为两种：一种是多级安全网络传输信道，另一种是普通网络传输信道。第 1 种网络信道中强调多级安全概念，这种信道存在于具有不同安全级别、需要进行隔离的主机之间。入侵者期望利用这种信道从高安全级别主机获得信息，并传递给低安全级别主机。第 2 种网络信道中并不局限于多级安全环境，普通网络中主机没有安全级别的定义。这种信道的两端主机甚至可能是被允许通信的。该信道只是期望在通信链路上再附加一层隐蔽通信。由于更贴近于日常应用，并且涉及更广泛的安全环境和策略，目前第 2 种网络信道的研究逐渐占据了主流地位。

1987 年，Girling 发现了 3 种局域网上的隐蔽信道，开启了对普通网络中隐蔽信道的研究。1996 年，Handel 对 OSI (Open Systems Interconnection) 网络模型进行了深入分析，提出了许多理论上的潜在隐蔽信道。同年，Rowland 在 TCP / IP 协议部分找到了许多隐蔽信道实例之后，网络信道的威胁得到了广泛的认识。网络隐蔽信道的识别、度量和处置也逐渐成为隐蔽信道研究领域的热点之一。

网络隐蔽信道也包括存储隐蔽信道和时间隐蔽信道两种类型。网络存储隐蔽信道主要是在各种协议的数据包中加载信息。为了实现隐蔽传输，一般将信息附加在不常用的数据段中，包括未用的 IP 头字段、IP 头的扩展和填充段、IP 标识和碎片偏移等。也有的网络存储隐蔽信道将信息隐藏在应用层编码中。网络时间隐蔽信道则一般利用网络中传输数据包的时间特性来表示信息，这些时间特性包括数据包的发送 / 到达时刻、间隔时间等。

4. 推理信道

推理信道一般存在数据库系统中，是指恶意用户利用历史访问记录查询信息，实现对敏感信息间接访问。在部分研究成果中，推理信道被归属于隐蔽信道领域。不过，这种信道与 TCSEC 标准中的隐蔽信道概念存在一些差距具体如下。

(1) 推理信道所针对的系统不只局限于多级安全系统，执行其他安全策略的系统中同样存在推理信道威胁。恶意用户可以利用推理信道以来授权的方式获取敏感信息。

(2) 推理信道中发送者并非是必须的，恶意用户可以独立构造查询来完成信息盗取。因此，推理信道并不需要植入木马。

6.1.4 隐蔽信道分析技术

隐蔽信道分析技术主要包括 4 个领域，分别关注隐蔽信道的建模、识别、度量和处置。其中，后三者被 TCSEC 标准规定为针对具体信道需要执行的工作。

信道建模关注于隐蔽信道产生原因的研究和信道的形式化模型表示，主要包括信息流模型和无干扰模型。

信道识别强调对系统设计和代码进行分析，寻找可能被用来构建信道、进行隐蔽通信的共享资源、原语等设施，即搜索 $\langle \text{variable}, \text{PA}_h, \text{PV}_i \rangle$ 中的 variable 变量和 PA_h 、 PV_i 原语。对系统执行信道识别操作，可以确定系统中是否存在隐蔽信道隐患。针对信道建模模型描述的隐蔽信道产生原因，分别形成了一系列隐蔽信道的识别方法，包括信息流分析方法、无干扰分析方法、共享资源矩阵方法、隐蔽流树方法、代码级分析技术、回

溯搜索法、逆向共享资源矩阵等方法。

在识别出隐蔽信道之后,系统的安全保障人员就可以对信道的传输能力及危害程度进行度量,并采取处置措施来保障系统的安全性。对隐蔽信道威胁的度量结果可以用来评价信道对系统安全性的威胁程度,并指导对其采取适当的限制措施,以有限的代价来保障系统的安全服务。TCSEC 标准中规定,使用信道容量作为信道威胁评价的指标。信道容量度量方法主要分为两种类型,分别是 Millen 提出的形式化方法和 Tsai 与 Gligor 提出的非形式化方法。除了通过精确计算或实验来获得信道的容量数值以外,实际应用中也可能只需要通过数学分析获得信道容量值的取值范围。除了信道容量之外,用来对信道威胁度量的指标还包括短消息指标、隐蔽信道因素、相对容量等。

信道的处置措施包括 3 类,分别是信道消除、限制、审计和检测。对隐蔽信道危害的消除措施包括修改系统、排除产生隐蔽信道的源头或者破坏信道的存在条件。但是信道消除的方法一般代价较高,不易实现。早在信息安全标准 TCSEC 中就已经承认了隐蔽信道难以消除的可能性,并认为当安全系统对隐蔽信道施加有效干扰后能够限制信道传输能力,从而确保恶意用户即使通过信道盗取了机密信息,也会因为信息的准确度有限或者数据已经过时而无法对系统安全构成威胁。因此,TCSEC 标准允许系统放弃完全消除信道,而选择限制信道能力的处置措施,将隐蔽信道的传输能力限制在不能有效传递信息、侵害系统安全的范围内。隐蔽信道限制措施的目标是破坏隐蔽信道的传输能力,可采取的操作包括添加干扰、添加延时等。

信道审计强调对隐蔽信道相关操作的监测和记录,而信道检测则是指从审计到的操作记录中筛选出实际使用的隐蔽信道记录。审计方法的关键是确定哪些事件和数据必须被记录,检测方法的关键在于如何区分实际使用信道时产生的记录和系统正常使用产生的记录。对信道使用情况的审计操作可以用来对入侵者进行威慑,以降低入侵者使用信道威胁系统安全的可能性。

隐蔽信道的识别是审计和检测信道使用状况的基础,而被标识的信道则是审计和检测的对象。利用信道检测结果,防御方能够获得更加明确的信道使用信息。在这些信息基础上,可以度量信道的实际传输能力,从而能够对入侵行为采取针对性措施。另外,从检测结果中也能够获得入侵者的信息,增加系统对入侵者的威慑性。

6.2 阈下信道

阈下信道是一类特殊的隐蔽信道,更接近于信息隐藏领域。下面首先介绍阈下信道的相关概念和存在性,然后介绍阈下信道模型和评价指标,最后介绍阈下信道构造方法。

6.2.1 阈下信道相关概念

1. 基本概念

阈下信道也称潜信道,这概念最早是由 Simmons 于 1978 年在美国圣地亚国家实验室提出的。之后 Simmons 做了较多的研究工作。1983 年 Simmons 通过一个在看守的完全监视下两个囚犯如何协商一个逃跑计划的例子引入了该信道并正式命名为阈下信道。它的定义可表述如下:所谓阈下信道是指在基于公钥密码技术的数字签名、认证等应用密码体制的输出密码数据中建立起来的一种隐蔽信道,除指定的接收者外,任何其他人均不知道密码数据中是否有阈下消息存在。阈下信道是信息隐藏技术中比较特别的一

类，从载体数据（密码数据）的性质到消息嵌入方法上都有其独特之处。目前，人们在信息隐藏技术中取得的进展，包括对信息隐藏系统的许多隐写分析成果，不能直接套用到阈下信道技术中来。

由于作为载体的密码数据具有不可修改性和不可伪造性（假设密码系统是安全的），阈下信道系统与一般的信息隐藏系统有很多不同的性质。图 6.5 是一个利用输出密码数据的某一固定位比特奇偶性所构造的 1 比特阈下信道实例。在图 6.5 中 k_i 为对应于比特 b_i 的一次一密隐写密钥，由收发双方所共享， K_p 和 K_s 为密码系统的公钥私钥对。设每一公开消息对应一个密文数据集，该集合可分成某固定位比特分别为奇数和偶数的两个集合，并分别表示消息比特“1”和“0”，嵌入时，根据由隐写密钥 k_i 加密的阈下比特值来选择相应子集中的元素输出密文。阈下收方根据收到密文的固定位比特的奇偶性提取阈下比特，再由 k_i 解出阈下消息 b_i 。而公开收方仍可对密码数据进行正常解密或验证以实现载体密码系统功能。由于对阈下消息进行了加密，即使有人怀疑阈下信道正被使用也无法检测出来。1 比特的阈下信道可以推广至多比特的阈下信道，这只需将前述的密文数据集分成相应的多个集合即可。

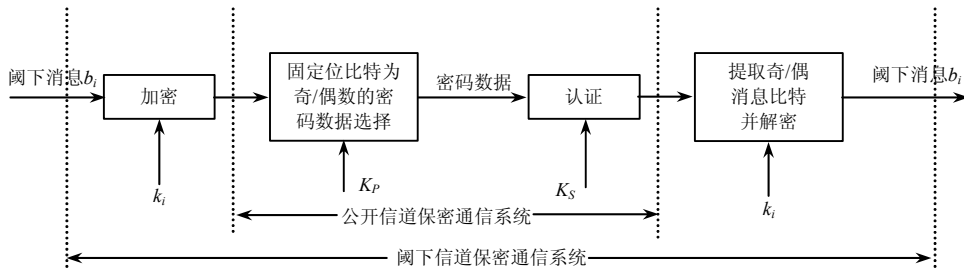


图 6.5 一比特阈下信道保密通信系统

2. 相关术语

阈下信道中的多数术语和一般信息隐藏技术是一致的，其中大部分在 1996 年的第一届国际信息隐藏会议中已达成了一致。以下我们对阈下信道系统中常用的和特有的一些术语作以简单的说明，还有些术语在后面提到时再给予说明。

(1) 阈下消息 (Subliminal Message)

通过阈下信道传送的秘密消息；

(2) 阈下发方 (Subliminal Sender)

阈下消息的发送者；

(3) 阈下收方 (Subliminal Receiver)

阈下消息的接收者，阈下收方与公开收方不必一致；

(4) 载体系统 (Carrier System)

在其中可以构造阈下信道的密码系统；

(5) 密码数据 (Crypto Data)

指密码系统的输出数据；

(6) 隐写密钥 (Stego Key)

用于秘密消息的嵌入和提取的密钥；

(7) 载体数据 (Cover Data)

是指一般信息隐藏系统中用于嵌入秘密消息的无害消息，在阈下信道系统中是指载体系统输出的密码数据；

(8) 隐写数据 (Stego Data)

嵌入秘密消息后的信息隐藏系统的输出数据, 或嵌入从阈下消息后的载体系统输出的密码数据;

(9) 无害消息 (Harmless Message)

看守允许传递的消息;

(10) 看守 (Warden)

检测阈下信道并阻止阈下通信的人。

6.2.2 阈下信道的存在性

阈下信道的载体系统是密码系统。就阈下信道存在的意义而言, 密码系统的输出数据不能有消息的保密性, 而只能有可验证性 (主要指认证性), 如果允许保密性存在, 则不如将阈下消息直接加密传输, 这失去了隐藏的意义^[93]。

一般地, 阈下消息与密码数据是相互独立的, 如果载体系统输出的密码数据是确定的, 即一个明文消息唯一对应一个密文数据, 则 n 比特阈下消息能够正确嵌入的概率仅为 $1/2^n$, 这样的阈下信道没有实用意义。如果一个明文消息有 m 个可能的密文数据输出且可从中随机选择一个作为密文, 那么当 $2^n \leq m$ 时任一 n 比特的阈下消息都可至少对应一个密文数据, 这时阈下消息可成功嵌入。可见, 密码系统的随机性是构造实用阈下信道的必备条件。

输出密码数据的随机性来自于系统的输入数据, 这里的随机性是针对看守而言的, 主要包括三类: 一类是系统输入的随机参数, 它是主要的随机源; 另一类是明文消息作为系统输入数据时其本身冗余度所引入的随机性, 即同一语义有多种可能的消息符号表示; 第三类是对于看守而言的发送消息的语义随机性, 即每次发送前看守对要发送的消息的不确知性, 不同语义之间如果其含糊度不为 0, 则相应的输出密码数据对看守来说就有随机性。其中含糊度意义如下: 设前 i 次通信的消息分别为 $m_1 \in M_1, m_2 \in M_2, \dots, m_i \in M_i$, 第 $i+1$ 次通信的消息为 $m_{i+1} \in M_{i+1}$, 那么前 i 次通信保留的第 $i+1$ 次通信的不确定度即条件熵 $H(M_{i+1} | M_1, M_2, \dots, M_i)$ 定义为含糊度。因此, 只要对看守而言密码数据中有前述的随机性存在, 就有阈下信道存在。

传统的对称密码体制, 包括分组密码和序列密码, 其本身就用于加密, 显然在这样的密码数据中建立阈下信道没有意义, 而且, 在确定的密钥下对称密码体制中的明文和密文是一一对应的, 通常没有数据扩展, 无随机冗余比特可利用, 所以对称体制中不存在实用的阈下信道。需要说明的是, 这里的随机冗余与纠错码的校验冗余是不同的, 在纠错码中消息比特唯一确定了冗余比特, 输出数据无随机性, 不能构造实用的阈下信道, 而随机冗余比特没有这样的确定性, 甚至和消息比特独立, 因此可以利用。

在非对称密码体制中, 比如签名、认证体制、零知识证明、秘密共享、电子现金等所采用的密码技术均具有概率特性, 一个明文 (或消息) 可对应多个密文。一个原因是某些密码体制其本身必须是概率体制才能保证安全性, 如 ELGamal 数字签名体制。另一个原因是在某些应用中, 明文空间很小, 如股票交易中的“买”和“卖”, 只有两个消息, 为了使消息能够准确认证, 必须采用签名等技术实现, 但如果密文空间很小, 即签名不是概率的, 则攻击者可通过简单的复制一条合法消息而达到买或卖的欺骗目的, 因此必须采用概率体制。这些概率特性的引入, 使得密码数据包含了较多的随机冗余数据, 这就为阈下信道的存在创造了条件。如果以阈下消息嵌入来代替部分或全部的随机冗余比特就构成了阈下信道。例如, 在 ELGamal 签名和 DSA 中, 签名的会话密钥是随

机选取的，在大多数签名中它也是必须的。因此一个消息在同一签名密钥下有若干个合法签名相对应，如果选取适当的会话密钥，使签名数据与阙下消息相适应，就既可实现阙下消息的传递，而又不影响公开收方的验证。

6.2.3 阙下信道的模型和评价指标

建立良好的阙下信道模型对于阙下信道的研究工作有着重要意义。本小节介绍阙下信道的一般模型和含有阙下信道的密码系统须满足的几个必要条件^[94]。

1. 阙下信道的一般模型

假定阙下消息的收发双方为 Alice 和 Bob，模型如图 6.6 所示。其中， $m \in M$ 是载体系统输入的原始消息； $u \in U$ 表示 Alice 要发给 Bob 的阙下消息； $u_r \in U_R$ 为随机化后的阙下消息，其中随机化密钥为 $k_1 \in K_1$ 。 $k_0 \in K_0$ 是 Alice 和 Bob 所共享的提取阙下消息所必需的陷门信息。对于有些阙下信道方案来说是不需要的，比如某些窄带阙下信道方案，此时记 $K_0 = \text{Null}$ （空集）。 k_0 和 k_1 合起来称为阙下信道的隐写密钥，记作 $k \in K = K_1 \cup K_0$ ； c 是载体系统的输出密码数据，如果系统有阙下消息嵌入，则称之为隐写数据记作 $c_s \in C$ 。否则称为原始密码数据记作 $c_0 \in C$ 。在不区分载体系统是否有阙下消息嵌入时统一记做 $c \in C$ ；嵌入算法 Emb 包括了两部分，一部分是载体系统的密码算法，它完成了载体系统的正常的密码功能，另一部分是阙下消息的嵌入算法，将消息嵌入于载体系统的输入或输出参数中；载体验证 Ver：是载体密码功能的实现过程，比如数字签名系统中的签名验证过程，同时也实现了 Bob 对隐写数据 c_s 的认证，即证明 c_s 确实来自发方 Alice；提取算法 Abs 是阙下收方 Bob 在载体验证过程完成后对阙下消息的提取过程。如果验证成功则 Bob 运行提取算法用隐写密钥 k_0 求解随机化的阙下消息 u_r ，如果 $u_r \notin U_R$ 则 Bob 可断定所收到的消息不正确，否则可由 k_1 恢复出阙下消息 u 。

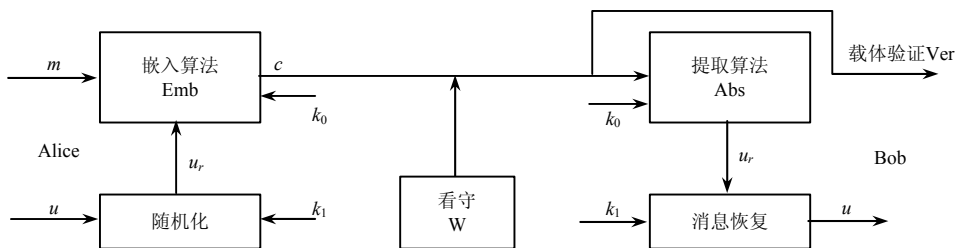


图 6.6 阙下信道的一般模型

下面给出含有阙下信道的载体系统须满足的几个必要条件，其中 Sys 为原始密码系统（即载体系统），Sys* 为嵌入阙下信道后的密码系统。

- (1) Sys* 的输入数据与 Sys 的输入数据的公开说明相一致；
- (2) 阙下消息嵌入算法包含于 Sys*，并且该嵌入算法是透明的，即它只影响 Sys 中的随机参数的选择，而不影响 Sys 中的算法描述的任何其他细节；
- (3) Sys* 的输出数据与 Sys 的输出数据的公开说明相一致；
- (4) 对除阙下收方以外的任何第三方，Sys* 保留了 Sys 的全部密码功能。
- (5) 载体密码系统每次所传输的公开数据应该是不引起任何第三方怀疑的无害消息，该消息不应具有保密性而只能有认证性（或可验证性）。因为如果允许具有保密性的数据传输，则不如直接加密传输，失去了隐藏的意义。同时任何第三方在不知道载体系

统私钥的情况下能够认证该消息,即他能够检验无害消息与相应的密码数据之间的关系。如果载体密码系统是安全的,则任何第三方都不能修改密码数据,否则就失去了消息的认证性,不能实现正常的密码功能。

目前,能够满足以上条件的密码系统主要有签名系统、某些认证系统、零知识证明系统、电子货币、电子护照、秘密分享和多方安全计算系统等。在这些系统中,看守可以对每一个数据进行正确的断言,即它是一个认证数据,签名数据或者是用于知识证明的数据等。在知识证明中,证明者的回答必须和验证者的询问一致才不会引起看守怀疑。

2. 容量

阈下信道的容量是指一次传输的密码数据所能携带的阈下比特数,由阈下信道存在性的讨论可知,阈下信道的容量与载体系统的输入数据的随机特性有关。把输入数据的集合记作 R_m ,它包括所有具有随机性的输入,如原始消息、秘密输入参数等。设载体系统的密码算法为 Cry ,则有

$$C=Cry(R_m) \quad (6.2)$$

实际上, Cry 可能包含有损变换,比如其中含有 Hash 函数,这会损失一部分输入数据的随机性,而输出集 C 包含了所有保留随机性,它体现了载体系统的潜在容量。然而阈下信道的实际容量是一个比较复杂的问题,受到诸多条件的限制,很难达到其潜在容量。

3. 嵌入空间

在嵌入过程中,随机化后的阈下消息 $u_r \in U_R$ 要编码成嵌入码字 $v \in V$ 。 V 是阈下消息要嵌入的载体系统输入参数集或输出密码数据集的子集,而输出是输入的函数,因此阈下消息的嵌入直接或间接地影响载体系统的输出数据的特性。我们把阈下消息嵌入的载体参数集统一记做嵌入空间 E ,则 $V \subseteq E$ 。

4. 安全度

系统需求指标是指一个系统为达到应用需求而必须满足的基本条件。比如在一般的信息隐藏系统中,主要原理是将消息隐藏于数字媒体的冗余当中,所形成的最终隐写数据将经过公开信道到达接收方。信道上存在各种敌手,他们的各种攻击行为对信道产生的影响构成了系统应用的实际环境,这些敌手分为两类:主动的和被动的。被动的敌手对消息进行侦听和检测,相应的信息隐藏系统提出了不可检测性需求指标,主动的敌手对传送的数据进行篡改,伪造和破坏,相应地又提出鲁棒性的需求指标。因此在这些系统中,抗攻击能力的研究就围绕着不可检测性和鲁棒性这两个指标来进行。

在阈下信道系统中,由于作为载体消息的密码数据的不可修改性和不可伪造性(假定载体系统是安全的),研究主要集中于抗击被动看守攻击的能力。因此,阈下信道所依托的载体系统的特殊性使得阈下信道的系统需求指标与一般的信息隐藏系统有所不同,这就是与不可检测性相对应的安全度。由前面介绍可知,看守是通过检测载体密码系统输出的随机性改变来判断是否有阈下信道存在的。秘密信道畸变度越小,实验统计分布接近原始分布的概率越大,可见该畸变度表示了阈下信道的安全程度,因此就把它定义为阈下信道的安全度。安全度的具体计算与攻击模型中畸变函数的选取有关,准确的说每一种选取都是安全度描述的一个具体模型。

5. 嵌入成功率

阈下消息在嵌入前必须正确编码为载体参数才行,然而并不总是成功的,这可能来源于以下几个原因。

(1) 作为载体系统而言, 每个秘密输入的随机参数只能使用一次, 否则将有可能泄漏系统的私钥。比如在 ElGamal 签名中, 如果两次用同一个会话密钥, 则任何第三方可解出签名私钥。举例如下: 假设消息 m_1 对应的签名为 (r_2, s_2) , 消息 m_2 对应的签名为 (r_2, s_2) , 如果两个签名所用的会话密钥均为 k , 则有 $r=r_1=r_2=g^k \bmod p$, 以及

$$s_1=k^{-1}(H(m_1)-xr) \pmod{p} \quad (6.3)$$

$$s_2=k^{-1}(H(m_2)-xr) \pmod{p} \quad (6.4)$$

联立方程 (6.3) 和 (6.4) 即可求得私钥 x 。因此, 如果两个阈下消息有相同的编码或者与用过的会话密钥相同, 都会威胁到密码系统以及阈下信道的安全性。如果阈下消息存在冗余度, 则可以改变其中一个消息的编码来避免以上问题, 否则以上情况一旦出现, 则不能成功嵌入。

(2) 在某些基于搜索的窄带信道中, 如果在搜索次数的最大上限之内没有找到合适的码字, 那么阈下消息就不能成功嵌入。

(3) 给定阈下信道方案 Sub, 对于阈下发方而言, 任意可能的阈下消息 $u \in U$ 的编码 $v \in V$ 必须满足一定的编码限制条件, 如果不能正确编码, 就成为不可嵌入消息。

综上, 阈下消息嵌入的成功率并不总是以概率 1 成立, 为了解决这一问题必须适当减少可嵌入容量。这说明嵌入成功率是影响容量的比较重要的参数, 我们就把阈下消息能够成功嵌入阈下信道并能够被收方正确提取的概率定义为嵌入成功率。

6. 信道使用率

当一个含有阈下信道的密码体制被多次使用时, 可能不是每一次都有阈下消息传递的, 这时如果不增加冗余度作为阈下消息的标记, 收方将无法判断提取的数据是否是阈下消息, 为此可以定义信道使用率参数, 记作 p_u 。如果一个信道在每 m 次使用中平均 n 次有阈下消息传递, 则该信道的使用率 p_u 定义为

$$p_u=n/m \quad (6.5)$$

在一个实用的信道中收方应能够以 1 的概率判定收到的数据是否是阈下消息, 如果 $p_u \neq 1$ 就必须通过在阈下消息中引入标识符来实现。比如当某一固定位比特值为“1”时可判定所提取的数据是阈下消息, 而“0”时不是。但这种策略要折衷安全度和信道容量。因为该固定位比特必然影响随机参数的概率分布, 同时也不能用于传递阈下消息。

6.2.4 阈下信道的构造方法

迄今为止, 人们已经构造了许多阈下信道, 在构造信道过程中所关心的问题除了载体系统本身的实用性之外, 主要集中于两个方面: ① 寻求新方法; ② 怎样在保证载体系统安全性的条件下逼近系统的最大容限。值得注意的是, 目前所构造的方案绝大多数为基于 DLP (离散对数困难问题) 或 ECDLP (椭圆曲线离散对数问题) 的密码系统, 且多数为数字签名系统。其主要原因是基于 DLP 的密码系统本身是概率的, 其数学结构非常有利于阈下信道的构造, 采用数字签名系统在构造难度和实用性等方面都有很大优势。根据阈下信道的不同构造机理, 阈下信道可以分成四类, 具体如下。

(1) 收发双方共享载体密码系统的部分或全部的陷门信息, 这些共享信息是收方提取阈下消息所必须的;

(2) 通过控制单向函数的输入或输出比特来嵌入阈下消息, 此时收发双方的计算复杂度是不对称的, 其中一方在提取或嵌入消息时要依赖于计算能力;

(3) 基于素性测试、产生元测试或可逆性测试等在公开密钥中嵌入阈下消息 (假设

载体系统的密钥对是由用户选取而不是系统分发的);

(4) 失败终止式阈下信道, 如消息嵌入成功则发送, 否则发方终止载体密码协议。下面首先介绍两种典型的签名方案: ElGamal 签名和 DSA 签名, 然后介绍上述四类阈下信道的构造方法。

1. ElGamal 签名与 DSA 签名

(1) ElGamal 签名

用户密钥的产生过程如下: p 是一个大素数, g 是 $GF(p)^*$ 上的一个生成元, 设用户的签名私钥为 $x \in_R \{1, 2, \dots, p-1\}$, 相应的公开密钥 $y = g^x \pmod{p}$ 。签名者 (Signer) 首先计算消息 m 的哈希值 $H(m)$, H 为标准的哈希 (Hash) 函数, 且有 $0 \leq H(m) < p-1$, 随机选择 $k \in_R \{1, 2, \dots, p-1\}$ 且满足 $\gcd(k, p-1)=1$ 。签名者计算

$$r = g^k \pmod{p} \quad (6.6)$$

$$s = k^{-1}(H(m) - xr) \pmod{p-1} \quad (6.7)$$

二元组 (s, r) 为有效签名。接收者通过检验 $1 \leq r < p$ 及 $g^{H(m)} = r^s y^r \pmod{p}$ 是否成立来验证签名。

(2) DSA 签名

用户密钥的产生如下: 大素数 $p \geq 512$ 比特, q 为 160 比特的素数, 且满足 $q|p-1$, $g \in GF(p)^*$ 是一个阶为 q 的元素, 设用户的签名私钥为 $x \in_R \{1, 2, \dots, q\}$, 相应的公开密钥 $y = g^x \pmod{p}$ 。对消息 m 的一个签名如下: 签名者首先计算 $H(m)$, H 为标准的哈希函数。然后签名者选取随机数 $k \in_R \{1, 2, \dots, q\}$, 并计算

$$r = (g^k \pmod{p}) \pmod{q} \quad (6.8)$$

$$s = k^{-1}(H(m) + xr) \pmod{q} \quad (6.9)$$

二元组 (s, r) 即为有效的签名。接收者通过检验 $r = (g^{(H(m)s-1) \pmod{q}} y^{(rs-1) \pmod{q}} \pmod{p}) \pmod{q}$ 是否成立来验证签名的正确性。

DSA 与 ElGamal 签名的显著区别是签名的生成子群不同, ElGamal 签名是在 $p-1$ 阶子群上生成的, 签名长度为 $2\lceil \log_2 p \rceil$ 比特, 而 DSA 签名是在 $p-1$ 阶群的一个 q 阶素子群上生成的, 签名长度为 $2\lceil \log_2 q \rceil$ 比特, 无论从签名长度和速度上 DSA 都有较大优势。

2. 第一类阈下信道

该类信道多数以 ElGamal 签名作为载体系统, 其中最经典的是牛顿信道 (Newton Channel), 是 Anderson 于 1996 年提出的 ElGamal 签名中的一种阈下信道, 利用 $GF(p)^*$ 中的平滑阶子群上离散对数问题的可解性来构造的。牛顿信道的具体方案如下。

令 $p=qt+1$, 其中 t 是 $p-1$ 的平滑因子。所谓“整数 t 是平滑的”是指对于一个适合的界 B , 若整数 t 的任何一个素因子 v 都满足 $v < B$, 则称 t 是 B -平滑的, 这里 B 是一个素阶群的最大可能阶, 在该群上的离散对数问题采用现有数学方法和计算能力是可解的。

阈下消息的嵌入分为以下步骤

(1) Alice 选取随机数 k' , 计算 $k = k't + u \pmod{p-1}$, 且满足 $\gcd(k, p-1)=1$;

(2) Alice 以 k 为会话密钥产生签名 (r, s) , 发送给阈下收方 Bob;

阈下消息的提取分为以下步骤

(3) Bob 首先验证签名 (r, s) , 如果不正确, 则拒绝该消息, 否则转到下一步;

(4) 首先看这样一个事实, 对 $r = g^k \pmod{p}$ 两边 q 次幂得到 $r^q = (g^k)^q = (g^q)^k = (g^q)^{k't+u} = (g^q)^u \pmod{p}$, 即提取阈下消息即是求解关于 z 的方程 $(g^q)^z = r^q \pmod{p}$ 。而 g^q 的阶 t 是 B -平滑的, 我们可以利用求解 DLP 的 Pollard- p 算法和 Pohlig-Hellman 分解算法求解 z , 运

算的时间复杂度为 $O(B^{0.5})$ 。

3. 第二类阈下信道

通过控制单向函数的输入和输出比特来嵌入阈下消息是一种常见的构造方法，这里单向函数的单项性依托载体系统所用的数学难题，它同时也是载体系统的安全基础，比如 ElGamal 签名中的 DLP 问题。下面介绍一种将消息嵌入于单向函数的输出数据的方案。在该方案中，阈下发方必须对单向函数的输入参数进行搜索以使输出与阈下消息相适应，我们以函数 Sch 表示搜索规则：以 N 表示搜索次数的最大上限；以 f 表示载体密码系统中的单向函数，如果输入为随机数 k ，则输出 $c=f(k)$ ； u 表示待嵌入的阈下消息； g 表示阈下消息的提取规则，在该规则下消息的提取是容易的， $u=g(c)$ 。载体系统仍以 ElGamal 签名为例，如图 6.7 所示。

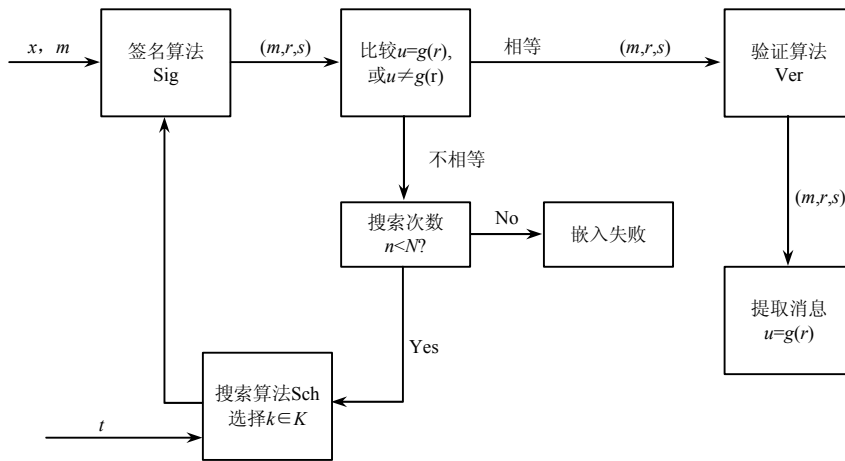


图 6.7 第二类阈下信道方案示意图

阈下消息的嵌入分为以下步骤

- (1) Alice 计算 $k=Sch(t)$ ，其中 t 为搜索函数的初始值；
- (2) Alice 以 k 为输入计算单项函数值 $r=f(k)=g^k(\text{mod } p)$ ；
- (3) Alice 比较 $g(r)$ 是否等于阈下消息 u ，如果相等则嵌入成功，以 k 为会话密钥产生 ElGamal 签名 (r, s) ，发送给 Bob；否则比较搜索次数 n 是否小于 N ，如果成立则转到 (1)，如果 n 大于或等于 N 则终止搜索，嵌入失败；

阈下消息的提取包括。

- (1) Bob 首先验证签名，如果不正确，则拒绝该消息，否则转到下一步；
- (2) Bob 利用规则 g 来提取阈下消息 $u=g(r)$ 。

由于选定单项函数输出值而求解输入值相当于求解 DLP 在计算上不可行，因此消息的嵌入必须依靠搜索合适的输入参数来实现。收方尽管能很容易地提取阈下消息，但不能求解单项函数的逆，从而保护了发方的签名权力。由于消息嵌入时必须采用搜索算法，因此可实现容量依赖于发方的计算能力。

第二类阈下信道的签名权力不能被收方滥用是一个优点，而付出的代价是容量受到了极大的影响，均系窄带信道，对其应用范围有一定影响。由单向函数的输入控制其输出特性需要很大的搜索量，控制选择的输出比特越多，需要的搜索量越大，计算复杂度越高。一般地，嵌入或提取的复杂度呈 2 的指数形式上升。

4. 第三类阈下信道

学者们提出的 SETUP 体制 (Secretly Embedded Trapdoor with Universal Protection) 主要研究了基于公钥密码体制中密钥对的一种消息泄漏方式, 相应的该类密码系统被称为受污染的密码系统, 它开发了载体系统的另一个随机源—初始化密钥。以下是基于 RSA 公钥的一个阈下信道典型构造实例。

假设 Alice 的 RSA 密钥对记为 $((n, e), (p, q, d))$, 其中 $n=pq$, $ed=1 \bmod \varphi(n)$, (n, e) 为公钥, (p, q, d) 为私钥, 为方便起见简记为 $((n, e), d)$, $\varphi(n)$ 为欧拉函数。Bob 的 RSA 密钥对记为 $((N, E), D)$ 。阈下消息为 u 。当 Alice 开始产生她的密钥对时执行以下协议。

阈下消息的嵌入包括。

(1) Alice 首先用 Bob 的公钥 (N, E) 对阈下消息 u 加密, 并将密文作为其加密指数 $e = \text{RSA}_{(N, E)}(u)$;

(2) Alice 随机选取 p, q , 并计算 $n=pq$;

(3) Alice 验证是否有 $\text{gcd}(e, \varphi(n))=1$ (元素可逆性) 且 $e < n$, 如果不成立则转到 (2), 否则继续;

(4) Alice 计算其解密指数 d , 完成其密钥产生并发布;

阈下消息的提取包括。

Bob 可通过简单地对加密指数 e 解密来恢复阈下消息 $u = \text{RSA}_{(D)}(e)$ 。

基于 SETUP 体制的阈下信道要求的条件很强, 载体系统的密钥对必须由用户随机选取而不是由系统分发, 而且必须在更换密钥时才能实行。一般的, 较长时间以后系统才更换一次密钥, 故使用效率很低。而且大多数的密码系统的密钥是由系统分发的, 因此这种阈下信道的实用性比较差。此外, 该阈下信道的发方安全性和第二类信道相同, 复杂度相对第二类小一些, 基本上与素性测试、产生元测试或可逆性测试的复杂度相当。

5. 第四类阈下信道

Desment 在对 Simmons 的封闭协议进行深入研究的基础上于 1996 年中引入了失败终止型 (Fail-Stop) 的阈下信道, 目的是为了指出 Simmons 的封闭协议并不能完全封闭阈下信道。该协议的思想是: 在嵌入过程中, 如果输出数据恰好与阈下比特相适应, 则嵌入成功, 阈下消息将被传递, 否则发方将终止协议, 但由于看守参与协议的执行, 终止协议将引起看守的怀疑, 因而有很大风险。该类信道没有实用性, 其存在价值是理论上的。

以上介绍的四类信道是从构造方法的角度来分类的, 值得注意的是它们几乎涵盖了所有可能的随机源的情况, 包括系统的各种输入参数 (公钥、协议运行中引入的随机数、公开输入的消息), 同时也对各种操作模式所能够引入的阈下信道进行了比较全面的总结。比如陷门信息泄漏模式, 搜索模式 (第二类信道), 基于素性测试, 产生元测试和元素可逆性测试等的在载体系统公钥中嵌入的模式以及失败终止模式等。

6.3 低截获概率通信

6.3.1 扩频通信技术

1. 扩频通信概述

扩展频谱通信简称扩频通信, 它是一种信息传输方式, 其信号所占有的频带宽度远大于所传信息必需的最小带宽; 频带扩展是通过一个独立的码序列来完成, 用编码及调制的方

法来实现，与所传信息数据无关；在接收端则用同样的码进行相关同步接收、解扩及恢复所传数据。扩频通信的射频信号频带宽度可扩展到信息信号频带宽度的数倍乃至数千倍。

扩频通信的基本工作原理：将发端输入的信息先调制形成数字信号，然后由扩频码发生器产生的扩频码序列去调制数字信号以展宽信号的频谱，展宽后的信号再调制到射频发送出去。在接收端收到的宽带射频信号，变频至中频，然后由本地产生的与发端相同的扩频码序列去相关解扩，再经信息解调，恢复成原始信息输出。用扩频函数调制和对信号相关处理是扩频通信有别于其他通信的两大特点。扩频通信工作方式有以下几种。

(1) **直接序列扩频** (DS-SS, Direct Sequence Spread Spectrum)。直接序列扩频是直接利用具有高码率的伪随机序列采用各种调制方式在发端扩展信号的频谱，而在收端用相同的伪随机序列去进行解码，把扩展宽的扩频信号还原成原始信息的扩频方式。

(2) **跳频扩频** (FH-SS, Frequency Hopping Spread Spectrum)。跳频的载频受一个伪随机码的控制，在其工作带宽范围内，其频率合成器按伪随机码的随机规律不断改变频率。在接收端，接收机的频率合成器受伪随机码控制，并保持与发射端的变化规律一致。跳频是载波频率在一定范围内不断跳变意义上的扩频，而不是对被传送信息进行扩频，不会得到直序扩频的处理增益。

(3) **跳时扩频** (TH-SS, Time Hopping Spread Spectrum)。跳时也可看成是一种时分系统，所不同的地方在于它不是在一帧中固定分配一定位置的时片，而是由扩频码序列控制的按一定规律跳变位置的时片。

(4) **脉冲线性扩频** (Chirp-SS, Chirp Spread Spectrum, 简称切普扩频)。发射的射频脉冲信号，在一个周期内，其载频的频率作线性变化。因其频率在较宽的频带内变化，信号的带宽也被展宽了。由于这种线性调频信号占用的频带宽度远大于信息带宽，所以也是一种扩频调制技术。

上面几种方式的时间和频率的关系如图 6.8 所示。扩频通信具有以下一系列优点。

(1) 抗干扰能力强。能在干扰环境中，通过分散功率或跳频等方式完成信息传输，达到抗干扰的目的。

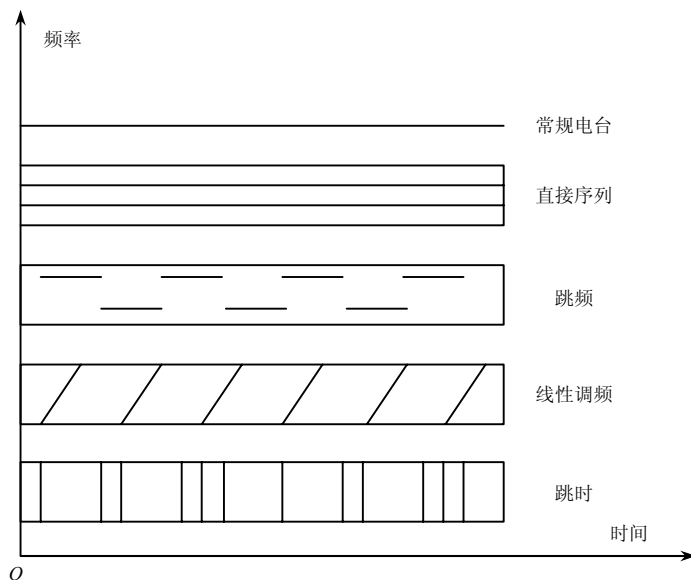


图 6.8 几种扩频方法之间的时间和频率关系

(2) 低截获率。直接序列扩频的射频信号功率分散, 淹没在噪声中; 跳频信号的频率在较宽的频带内跳变, 不易被敌方截获。

(3) 用作码分多址通信。在通信网中, 采用不同的码序列作为地址码, 发信端根据接收端的地址码选择通信对象。

(4) 抗多径干扰。在无线电通信的各个频段, 短波、超短波、微波和光波中存在大量的多径干扰。一般采用分集接收技术, 或设法把不同路径的不同延迟信号在接收端从时间上对齐相加, 合并成较强的有用信号, 这两种方法在扩频通信中很容易实现。

(5) 适合数字话音和数据传输, 以及开展多种通信业务。扩频通信一般都采用数字通信、码分多址技术, 适用于计算机网络, 适合于数据和图像传输。

(6) 安装简便、易于维护。扩频通信设备是高度集成, 采用了现代电子科技的尖端技术。因此, 十分可靠、小巧, 大量运用后成本低, 安装便捷, 易于推广应用。

(7) 有一定的保密性。尤其是跳频通信以其良好的抗干扰能力和多址性能引起了人们的很大重视, 目前正处在大量涌向军事用户市场的浪潮上。

2. 直接序列扩频通信

直接序列扩频(简称直扩)工作方式是目前应用广泛的一种扩频通信方式。直扩通信同其他扩频工作方式比较, 实现频谱扩展方便, 无论对通信、测距或是对其他应用都很合适, 是目前应用的最多、也是最典型的一种。直扩系统的结构框图如图 6.9 所示。输入的数据信息 D , 经过信息调制后变成带宽为 B_1 的调频(FM)或调相(PM)信号, 再由伪随机编码(PN 码)调制成带宽为 B_2 的宽带信号发射。接收机接收到信号后, 首先通过同步电路捕获发送来的 PN 码的准确相位, 由此产生与发送来的伪随机编码相位完全一致的接收用 PN 码, 作为扩频解调用的本地信号, 以便准确恢复成窄带信号, 从而获得对发送来的信息数据 D 的估计值 D_1 完成接收。

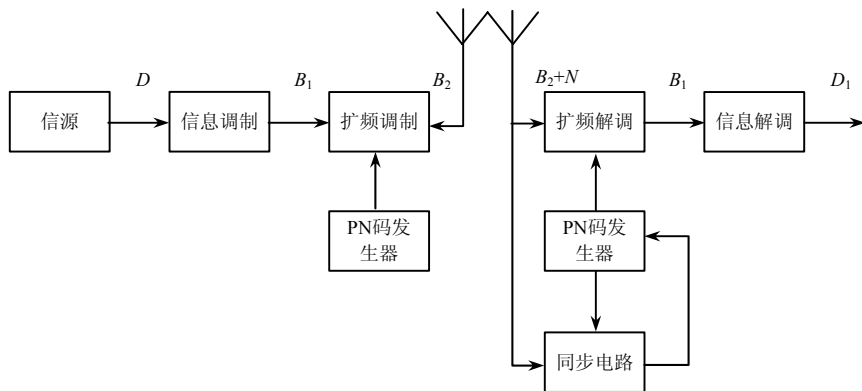


图 6.9 直扩系统结构框图

直扩系统的特点有以下几点：① 频谱的扩展是直接由高码率的扩频码序列进行调制而得到的；② 扩频码序列多采用伪随机码，也称为伪噪声(PN)码序列；③ 扩频调制方式多采用 BPSK(Binary Phase Shift Keying)或 QPSK(Quadrature Phase Shift Keying)调制。扩频和解扩的调制解调器多采用平衡调制器，制作简单又能抑制载波；④ 模拟信息调制多采用调频(Frequency Modulation, FM)，而数字信息调制多采用脉冲编码调制(PCM)或增量调制(Delta Modulation, DM)；⑤ 接收端多采用产生本地伪随机码序列对接收信号进行相关解调，或采用匹配滤波器来解扩信号；⑥ 扩频和解扩

的伪随机码序列应有严格同步，码的搜捕和追踪多采用匹配滤波器或利用伪随机码的优良相关性在延迟锁环中实现。⑦ 一般需要窄带通滤波器来排除干扰，以实现其抗干扰能力的提高。

在扩频通信系统中，信号频谱的扩展是通过扩频码实现的。扩频系统的性能和扩频码的性能有很大关系。对扩频码，通常有下列的要求：① 易于产生；② 具有随机性；③ 具有尽可能长的周期，使干扰者难以从扩频码的一小段去重建整个码序列；④ 具有双值自相关函数和良好的互相关特性，以有利于接收时的截获和跟踪，以及多用户应用。从理论上说，用纯随机序列去扩展信号频谱是最理想的。但在接收机中为了解扩应当有一个同发射端扩频码同步的副本。因此，在实际中，我们只能用伪随机或伪噪声(PN)序列作扩频码。

伪随机序列具有类似噪声的性质，但它又是周期的有规律的，既易于产生，又可以加工和复制的序列。伪随机序列应当具有类似随机序列的性质，而随机序列具有的性质归纳起来有以下三点。

(1) 随机序列中的 0 和 1 的个数接近相等。

(2) 把随机序列中连续出现 0 或 1 的子序列称为游程，连续的 0 或 1 的个数称为游程长度。随机序列长度为 1 的游程约占游程总数的 $1/2$ ，长度为 2 的游程约占游程总数的 $1/2^2$ ，长度为 3 的游程约占游程总数的 $1/2^3$ ……在同长度的游程中，0 游程数和 1 游程数大致相等。

(3) 随机序列的自相关函数具有类似于白噪声自相关函数的性质。伪随机序列具有类似随机序列的性质，但它的结构或形式预先可以确定，并且可以重复的产生和复制。

在直扩系统中，常见的扩频码有 m 序列、 M 序列、Gold 序列等。这几种序列都具有前面提到的伪随机码特性。而由于 m 序列易于产生、规律性强、有许多优良的特性，在各种扩频码中最早得到广泛应用，且理论研究最深入。 m 序列，是最长线性移位寄存器序列的简称。顾名思义，它是由多级移位寄存器或其他延迟元件通过线性反馈产生的最长的码序列。线性反馈移位寄存器的结构图如图 6.10 所示。在二进制移位寄存器中，若为 n 级数，则所能产生的最大长度的码序列为 $2^n - 1$ 位。因此， m 序列的最大长度决定于移位寄存器的级数，而码的结构决定于反馈抽头的位置和数量。

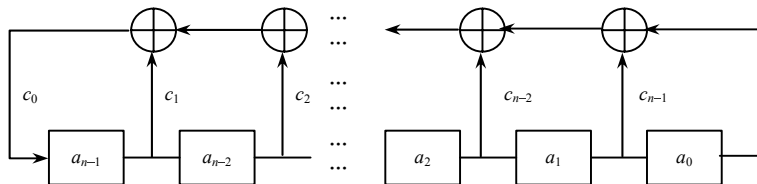


图 6.10 线性反馈移位寄存器

m 序列有下列一些基本性质：① m 序列中一个周期内 1 的数目比 0 的数目多 1 位；② 一般来说， m 序列中长为 R ($1 \leq R \leq n-2$) 的游程数占游程总数的 $1/2^R$ ；③ m 序列具有双值自相关函数特性，即 m 序列的自相关函数在原点处取值为 1，其他位置为 $-1/(2^n - 1)$ ；④ m 序列和其移位后的序列逐位模 2 相加，所得的序列还是 m 序列，只是相移不同而已；⑤ 序列发生器中移位寄存器的各种状态，除全 0 状态外，其他状态只在 m 序列中出现一次。由于 m 序列具有均衡性、游程的分布、自相关特性与随机序列的基本性质相同，且 m 序列具有一定的随机性和一定的周期性，故它是一种伪随机序列。

直扩通信系统的扩频调制主要有 BPSK, QPSK, MSK (Minimum Shift Keying) 等方式, 而 BPSK 调制是采用较广泛的一种调制方式。BPSK 调制即 180° 二相相移键控调制, 在数学上它可以用载波和一个取值为 ± 1 的伪随机序列 $C(t)$ 函数的乘积来表示, 调制框图如图 6.11 (a) 所示。若二进制信息为 $d(t)$, 取值为 ± 1 , 伪随机序列 $C(t)$ 的码元宽度为 T , 序列长度为 p , 带宽为 B_c 的高速二进制信息, 取值为 ± 1 。信息数据 $d(t)$ 调制伪随机序列 $C(t)$ 的模 2 加法器由简单的异或门实现, 已调序列 $d(t)C(t)$ 具有与伪随机序列 $C(t)$ 相同的功率谱密度。载波调制器是一个模拟乘法器, 对已调序列 $d(t)C(t)$ 与载波相乘, 即得二次调制后的发射信号

$$S(t) = \sqrt{2P}d(t)C(t)\cos(\omega_0 t + \psi) \quad (6.10)$$

式中, P 为恒包络数据调制载波功率; ω_0 为角频率, ψ 为随机相位。在发射端, 信息数据 $d(t)$ 也可以先进行载波调制, 再进行扩频调制, 调制框图如图 6.11 (b) 所示。

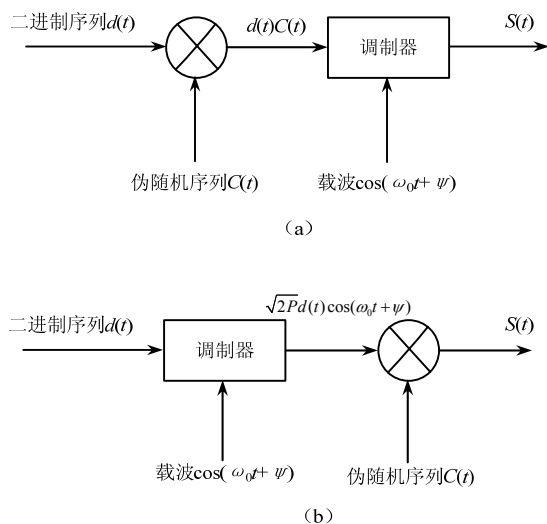


图 6.11 BPSK 直扩通信系统发射电路

3. 跳频通信

和其他几种扩频方式相比较, 跳频具有良好的抗远近效应的性能, 而且实现起来比较简单, 同时有同步捕获时间短等优点, 这使得跳频技术成为战术通信的首选抗干扰技术。目前大多数战术电台使用的都是跳频通信方式。所谓跳频, 用简单的术语表达就是“多频、选码、移频键控”, 即用伪随机码序列构成跳频指令来控制频率合成器, 并在多个频率中进行选择的移频键控。我们熟悉的二元频移键控 (Binary Frequency Shift Keying, 2FSK) 只有两个频率 f_1 和 f_2 分别代表传号和空号。而跳频系统则要求提供几百个, 甚至上万个频率。现在, 已经有实际的系统具有 2^{20} 个离散的频率供随机选择。由所传信息码与伪随机序列模 2 加 (或波形相乘) 的组合来构成跳频指令 (又称跳频图案), 并由它来选择发送频率。与定频无线通信相比较, 跳频无线通信对干扰是躲避式的, 可以避免多径干扰和瞄准式干扰, 同时具有对抗敌方侦察接收机截获破译的能力。

然而, 在现有跳频通信系统中, 通常只采用了一种调制方式。截获方一旦得到通信方的跳频图案, 可以很容易实现解跳, 然后采用调制识别技术就能够得到信息发送方的信息, 这样非常不利于信息传输的安全性。目前, 国内外对跳频信号的截获接收技术研究已经比较成熟, 跳频图案的截获、跟踪不再是困难的事。另外随着调制识别技术研究的不断深

入, 实用的调制识别方法层出不穷, 固定的调制方式很容易被对方识别从而解调出信息。

如果在跳频通信中不再采用固定的调制方式, 而采用调制跳变的方法, 这样当跳频图案被破译后, 还可以通过调制跳变来保障信息的安全性。理论和实践已经证明, 在同等条件下, 对同一种调制方式的信号的时间观测样本越多, 调制识别正确的概率越高。另外, 现有的调制识别技术通常针对的是某一类或某几种调制方式所作的识别工作, 当面临的调制集合不再固定或是跳变的时候, 现有的调制识别方法将会失效。

在跳频通信中采用调制跳变技术, 可以使不同载波频率上的调制不相同, 信道上通信信号的特征不再是恒定不变的, 即使对方能够正确解跳并进行信号拼接, 但对调制跳变信号却难以进行调制识别。因此, 在跳频通信中采用调制跳变技术对于提高信息传输的安全性和低截获性具有积极的意义。

4. 跳端口通信

由于计算机系统及网络技术本身存在着脆弱性、缺少安全性防护措施、缺乏安全性实践等, 在这种情况下, 在计算机网络中的数据传输可能遭遇到各类攻击如 DOS (Denial of Service)、信息流重定向 (流入到错误的地址或端口)、信息被窃取、数据畸变 (不完整性) 等。这些问题给计算机网络信息安全造成了重大威胁, 尤其是在军事领域, 通信安全问题显得尤为重要。通过借鉴调频通信的思想, 美军陆军研究实验室 (ARL) 正在研究“跳端口 (Port Hopping, PH)”技术, 它是针对 Internet 的安全问题提出来的。

传统的通信方式是收发双方约定一对固定的端口号, 这十分类似无线通信中的定频通信。攻击者只要掌握了通信双方的端口号或中心频率, 就能全部截获通信的内容。而跳频 (Frequency Hopping, FH) 通信时双方拥有相同的频点资源和跳频图案, 通信时双方实际使用的频率变化可以高达每秒几百次到上千次, 第三方企图通过同步跟踪的方法窃取信息是难以实现的, 只能采用全频段阻塞干扰的方式加以破坏。PH 利用了扩频通信中跳频的思想, 使信息传输通过跳端口的形式进行。PH 技术可以是基于数据包或是基于会话, 每一个数据包和每一次会话所使用的端口号都是不一样的。发端通过组合数据包或会话, 并以通过跳端口的形式进行传输, 就像跳频通信一样, 把信息隐藏在“Internet 噪声”之中。该技术总共可以使用 64K 个随机的端口来传输会话或数据包, 从而大大地减少了敌方窃取信息的能力。由于收端和发端知道具体的跳端口方式, 所以它们之间可以进行可靠的通信。

实现 PH 技术对网络的同步要求更高, 因为发端和收端要同时清楚地知道数据包的跳端口图案, 即下一个数据包或下一次会话所用的端口号, 否则就会因为不同步而丢掉数据包, 从而达不到跳端口目的, 甚至连正常的通信都无法保证。

6.3.2 流星余迹猝发通信技术

猝发通信是现代通信技术中具有很强抗干扰的一种通信方式。由于采取突发方式传送信息, 传输时间是随机的, 持续时间非常短, 在瞬间将信号传送完毕, 峰值功率很大。因此它具有隐蔽保密、抗截听、抗侦查、抗干扰等优点, 特别适合战地通信。

流星余迹猝发通信是利用流星电离余迹对 VHF (Very High Frequency) 无线电波的反射和散射作用来进行通信的。流星碎片从外太空进入大气层时发生灼烧, 能在离地面 80~100km 的地方形成 100km 左右长的圆柱体电离带。它能维持几秒钟内 2000km 以下的无线通信。流星余迹电离通信的频率范围在 30~100MHz, 最佳通信频带为 40~50MHz。它必须建立一个主站, 用于发射信号探测是否有合适的流星余迹可用于通信。

传统的流星猝发通信信息发射协议是：主站或从站将要发射的信号存储好，等待合适的流星余迹，当有合适的流星余迹通过所需空域时，从站会将响应消息和存储的信息一起发射出去。如果该响应到达主站则意味着此时通信链路是开通可用的，从而数据包会从主站发送到从站。流星电离余迹的可用时间一般很短，仅为几百毫秒到 1s。因此一个余迹消失后，要等待下一个适用的流星余迹出现，其等待时间一般为几秒钟至几分钟，甚至更长。显然，这种通信方式只能是间断的、突发的，故也叫流星余迹猝发通信。

1. 流星余迹的猝发通信机制

按照流星余迹的电子线密度，流星余迹可分为两类：过密类和欠密类。其中，电子线密度大于或等于 $2 \times 10^{14} \text{e} / \text{m}^3$ 的称为过密类，小于 $2 \times 10^{14} \text{e} / \text{m}^3$ 的称为欠密类。随着余迹的扩散，电子线密度随之下降。对欠密类余迹来讲，从流星余迹反射的信号幅值随时间按指数规律下降，余迹反射的信号功率为

$$P_{\text{rec}}(t) = p_0 q^2 e^{-2t/\tau} \quad (6.11)$$

式中， p_0 是 $t=0$ 时接收端接收到的每单位电子线度上的功率； q 为余迹的电子线密度； τ 为衰减指数，它与工作波长、扩散系数和电波入射角有关。

对于欠密类流星余迹，接收信号功率随时间呈指数快速衰减，使得一次流星余迹的可用时间很短，且通信方式也只能是间断式的。在这种通信机制下，要获得较高的数据通过率，必须使用具有较高速率的数传方式。对于过密类余迹，其通信时间也很短，并且在一小段时间内其衰减是很快的，但信号幅度在整个范围内并不是单调递减的而是震荡衰减的。国内外专家在提高流星余迹通信的传输速率时，发现实际信道容量的提高远小于理论预计值，这是由于多径传播引起的。因此抗多径干扰也是流星余迹通信要解决的一个重要问题。

2. 流星猝发通信系统的工作模式

流星余迹猝发系统主要由三种工作模式：点对点通信、网络通信和广播。当系统处于工作状态时，所有的猝发通信装置将进行点对点通信。猝发通信在这三种模式下猝发通信将如何进行系统配置呢？

点对点工作模式，其工作是比较直接的，它只需要系统具有发射信号能够被终端设备精确的接收，保证猝发通信的开始与结束均处于有效期。对发射机而言开始太晚或结束太早，对有效猝发通信时间的一种浪费。如果发射时间超过猝发通信范围将导致高误信率。在所有的点对点系统中，存在反馈路径。在半双工模式中，信号向前传播和信号反馈将共享相同的频率，但是全双工则用的是不同的频率。为了达到在 A、B 两个不同点直接进行数据的发送，每个发射机都用他们的各自不同的频率，同时主站（这里指 A 点）不停地发射探测信号。当 B 点监测到这个探测信号时，它发射一个引导信号以及它自己的信息。A 点用引导信号来同步它的接收机，随后它就能接收数据了。早期的点对点猝发通信系统用信号的门限来决定探测信号是否被收到。后来 JANET 系统发展到用信噪比来替代门限检测，降低了虚警率。再后来发展到 JANET B 系统时，SHAPE 技术中心考虑到信道利用率与低误信率之间存在着不可调和的矛盾，即使用信噪比也无法解。所以他们又利用 ARQ (Automatic Repeat reQuest) 拍发每 7bit 为一个数据包的电传。这种方法后来被用到 COMET 系统中。

流星余迹猝发通信在网络通信方面并没有进行很好的探究，西方硬件联盟为农业部安装的网络操作系统，它由 511 个遥控站，两个主控站组成，同时系统配置了双向数据传输功能。通过由一个主控站实行存储前向工作模式来实现网络通信。

6.4 匿名通信

6.4.1 匿名通信概述

网络作为通信和信息发布的工具被越来越广泛地使用,用户的各种需求也随之受到越来越多关注,特别是在网络使用过程中个人隐私及信息通信安全等需求。由于网络的开放性,人们能在更大的范围内交流信息、共享信息。但也正是因为网络的开放性,恶意节点也可加入,而又由于恶意节点行为的不可预测性,用户越来越多的信息受到非法窃取的威胁。不但是传输文件的内容,攻击者还能通过传输数据流的分析获知有关文件发送者与接收者身份等信息。对于很多网络应用而言,包括电子投票系统、电子商务、电子拍卖系统、特别是一些军事、国防部门的通信,如何保护隐私信息和秘密信息不被泄露以及如何保证网络通信的安全已成为当前的最基本要求。

广义的网络信息通信安全不仅是指保证信息内容的私密性、完整性、可用性、可追溯性和不可否认性,还涉及与信息流相关的通信连接安全性。从信息完整性与可用性来说,目前已有各种比较成熟的数据加密技术来保证,这些数据加密技术通过对信息内容进行加密,达到隐藏网络通信中信息内容的目的,可较好地保证网络传输信息的安全性。但是,数据加密技术对信息私密性以及通信连接安全性保护是不够的,它们不能隐藏 TCP / IP 等协议中报文的头部信息,如源地址、目的地址、报文长度等,即不能隐藏有关通信中发送者或接收者的位置信息和通信模式。攻击者可利用协议所存在的漏洞发动窃听与流量分析攻击,获取通信双方地址,推断出一些有价值的信息,包括通信双方的地址、谁和谁在什么时间通信、哪些 IP 在使用哪些服务等。这些都将给通信双方的安全隐私带来威胁,因此仅从数据加密的角度保证网络信息通信安全是不完整的,研究实现信息私密性保护、防止通信双方隐私信息泄露的防御性安全技术是当前保证网络信息通信安全必不可少的部分。

匿名通信技术作为一种防御性安全保护手段,逐渐成为网络安全研究领域的热点,受到越来越多网络安全研究人员的重视。匿名通信技术是指通过匿名计算方法将网络负载中的通信关系进行加密,使窃听者无法推知网络通信实体的具体位置或隐私信息,从而保持网络通信行为的不可观测性和隐蔽性,实现对网络应用通信实体的隐私及涉密通信更好的保护。概括来说,包括 Internet 网络中匿名 Email、Web 消息浏览与发布、电子支付等各种应用以及军事、国防等移动通信与计算在内的很多网络应用程序都有匿名的需求,都需要利用匿名技术来实现用户身份与节点信息的隐蔽性以及数据传输的隐蔽性。随着日益广泛的网络应用中对个人隐私及信息安全的要求提高,以及军事、国防、政府等国家部门对涉密通信的保护要求提高,研究匿名通信技术以及匿名通信技术在不同应用目的和环境中更好地实现都具有非常重要的理论意义和应用价值。

自从 1981 年 David Chaum 提出匿名的概念以来,对于匿名通信技术及原型系统的研究就受到了研究者的重视。到目前为止,已经有很多匿名通信系统及匿名通信技术的理论和实现。目前的匿名通信系统,根据所要隐藏信息的不同,有三种匿名保护形式:**发送方匿名**(Sender Anonymity)保护通信发起者;**接收方匿名**(Recipient Anonymity)保护通信接收者;**收发双方无关联**(Unlinkability of Sender and Recipient)保护发送者与接收者无法被关联。发送方匿名是指无论是接收者还是攻击者,都无法从通信流的内容或形式上,找出通信流的真实发送者。发送者匿名主要通过广播形式或者通过选择多级目标来实现。接收者匿名是指即使在接收者可以获知发送者身份的情况下,发送者和攻击者

也不能确定通信流的接收者。收发双方无关联是指攻击者在获知了若干个发送者组成的集合和若干个接收者组成的集合正在通信,但却无法将两个集合间的成员一一对应。

目前的匿名通信协议和技术是针对各种不同的环境下提出来的,且各自的实现机制不同,但总的思想是一样的,它们可以抽象为如图 6.12 所示的基本通信模型。在模型中,通信的请求方 I (Initiator) 是产生数据信息 V 并请求建立匿名通信的用户;响应方 R (Responder) 是与请求方 I 进行通信并接收数据信息 V 的一方;信息延迟代理方 A (Agent) 是用户可信赖的中间机构,它负责进行匿名通信,以保护用户的信息不泄露。通信双方各有一个代理,分别标记为请求方代理 A_I 和响应方代理 A_R 。根据报文信息的发送和接收过程,报文可以分为以下五类: I-A 型表示报文从请求方 I 到代理 A_I ; A-A 型表示信息从一个代理到另一个代理; A-R 型表示信息从代理方 A_R 到响应方 R; R-A 型表示应答信息从响应方 R 到代理 A_R ; A-I 型表示信息从代理 A_I 到请求方 I。

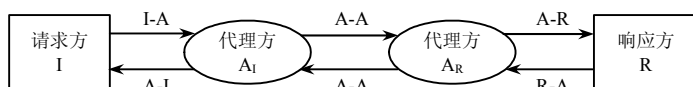


图 6.12 匿名通信的基本模型

下面概括介绍一些典型匿名通信系统体系结构^[95],以及评价指标—匿名性能与效率。

6.4.2 匿名通信系统体系结构

匿名通信系统体系结构主要是指匿名通信系统的具体应用实现及普适性问题,研究者们已经开发出很多原型系统及通信协议,目前,大多是针对具体应用环境进行设计的,如针对有线网的 DC-Nets、洋葱路由和 Crowds 以及针对移动自组网的 ANODR、MASK 协议等。这些匿名通信协议和系统能提供发送者匿名、或接收者匿名、或通信关系的匿名、或这三者的组合。尽管这些匿名协议与系统在目标上有共同之处,但各自为体系,缺乏通用性,从匿名通信系统结构考虑,可以将各种原型系统和协议分为有线网体系结构和无线网体系结构两大类,有线网体系结构分别为基于广播/组播机制的匿名系统体系结构及基于转发机制的匿名系统体系结构。

1. 基于广播/组播的匿名通信系统

基于广播/组播的匿名通信系统主要通过网络的广播或组播机制达到隐藏发送者或接收者的目的。广播与组播通信可完成主机间一对多的通信,基于广播/组播的匿名通信系统结构中,发送者节点利用广播或组播技术将信息传送到包含接收者的一组节点中。广播组或组播组成员越多,攻击者能猜中发送者或接收者的概率越小,匿名性就越好。典型的系统包括有 DC-Nets、P5、Hordes 和 Mapper 等。

DC-Nets 采用时间轮限定一个发送者在一轮内广播信息,而其他参与者在该轮内广播噪音来实现发送者匿名,该协议是唯一一个在缺乏可信机构的前提下,提供可证明的安全性的系统。P5 是一个基于层次广播结构和公钥加密机制的非中心化匿名通信协议,在不同的层次中 P5 提供不同等级的匿名性,所有成员沿建立好的逻辑广播二叉树结构向上或向下广播有用信息或垃圾信息,从而能够实现发送者匿名、接收者匿名以及发送者、接收者匿名;其强匿名性及良好的可扩展性是通过牺牲系统效率而得到的。Hordes 利用组播通信的方式来实现接收者匿名,当接收者返回消息应答时,采用组播通信来实现匿名连接,这样,即可以通过回复数据包中的组播地址隐藏发送者的目的,同时可以

减少匿名通信参与者所需工作量,降低数据传输延迟和数据链路的占用率;与之相类似的机制还有 DMCA / SMCA、SAM、Mapper 等,这些机制都考虑了组播的特性来获取匿名性能。

基于广播 / 组播的匿名通信系统利用广播 / 组播特性使参与者的匿名性能得到保证,但是,也正是由于其广播 / 组播特性,基于广播 / 组播的匿名通信系统的效率普遍较低,且其可扩展性较差,并不太很适合大规模动态网络中。

2. 基于转发机制的匿名系统体系结构

基于转发机制的匿名系统体系结构是目前应用最多最广的一种体系结构,大部分匿名系统都是基于转发机制的。它也可以称之为基于代理的重路由匿名通信系统,主要是利用数据包在网络中节点的转发来达到匿名的目的。基于转发机制的匿名通信系统可以分为基于简单代理的结构和基于多代理的结构,其典型的代表包括最早出现的匿名通信系统 Mix、基于单代理的 Anonymizer、基于组群的 Crowds 等。

Anonymizer 是一个基于简单代理的、提供匿名 Web 访问服务的匿名通信系统结构,在系统中需要匿名的用户通过代理的转发来进行 Web 访问,由此,Web 服务器只知道代理的存在而无法发现真正通信的发送者,从而达到隐藏发送者的目的。基于简单代理的匿名通信结构能实现发送者匿名,适合于匿名 Web 访问的应用服务,但是,由于只有一个简单的代理服务器或单个的匿名服务器,当其被攻击者所攻陷,所提供的匿名性能将变为 0,因此,简单代理所提供的匿名是有限的。

Mix-Net 是第一个源路由结构匿名通信模型,通过专用的 Mix 集群和加密、混淆、排序等技术,使经由 Mix 出入的消息包达到不可链接,并隐藏用户身份信息。基于 Mix 机制上,研究者还研究了多种 Mix 代理上数据包缓冲处理的机制,这些机制与 Mix-Net 致力于提供发送者匿名、接收者匿名以及发送者和接收者的不可链接性,但是,由于在每个 Mix 节点上要经过数据缓冲处理,造成匿名通信延迟长,因此,基于 Mix-Net 机制的匿名通信系统只适用于高延迟网络中。Tarzan、Onion Router 等也是基于 Mix 思想的匿名通信系统,其出发点是改善 Mix 机制带来的通信延迟代价,建立适合低延迟网络的匿名通信。Tarzan 用 P2P 网络与 Mix 机制,引入 IP 隧道技术和覆盖流技术、嵌套加密方法和多跳转发路由实现匿名;Onion Router 利用 3 跳的洋葱嵌套加密机制实现发送者匿名和接收者匿名;Tor 将 Tarzan 的基于线路的低延时思想与 Onion Router 嵌套加密机制相结合,从而增加转发的秘密性、拥塞控制、目录服务、完整性检测及退出策略的可配置性。这些匿名通信系统能提供实时匿名服务,如匿名 Web 访问服务等,但这类匿名通信系统由源节点封装数据包,在通信前必须选定路径上的转发节点,确定好转发路径。

这里稍微介绍一下**洋葱路由**(Onion Router, OR)技术。它是美国海军研究实验室的研究者们采用多次混淆的办法提出的一种匿名通信技术。OR 是为了阻止在公用网络上进行窃听和流量分析,以提供双向、实时的匿名连接,可以在公开的计算机网络中隐蔽网络的结构,对在互联网上进行的跟踪、窃听和流量分析有很强的抵抗作用,通信双方用洋葱包代替通常的 TCP/IP 数据包,利用代理技术实现与目标系统间的连接。这种洋葱路由代理技术的实现过程如图 6.13 所示。OR 通信过程可以简单地描述如下:当主机 A 要与主机 B 进行通信时,主机 A 向代理路由器 W 发送一个数据包,代理路由器对该数据进行封装,封装是按分层进行,首先是对代理路由器 Z 的地址和要发送的数据进行封装,把刚封装好的包与路由器 Y 的地址再进行二次封装,然后再把所得的包与路由器 X 的地址进行封装,这样封装好的数据包类似于洋葱的结构,故称为洋葱包。当代理路由

器 W 把这个洋葱包传到路由器 X 时, 路由器 X 用自己的密钥对所得的数据包进行解密, 得出下一站是路由器 Y, 于是路由器 X 就剥掉包中指明 Y 的地址及包头信息, 并对洋葱包进行填充, 使洋葱包的大小不变, 然后把处理后的洋葱包传到下一个路由器 Y, 路由器 Y 收到了洋葱包后, 按照同样的原理把洋葱包又给下一个洋葱路由器, 最终传到终点路由器, 即代理路由器 Z 用自己的密钥解密洋葱包, 得到主机 B 的地址后, 直接把数据包发送给主机 B。窃听者如果监听链路中的某个节点, 如路由器 Y, 只能得到洋葱包是从路由器 X 发来的以及下一站是路由器 Z 的信息, 不可能得到其他任何信息。为了防止流量分析, 在节点转发时可以加入填充技术, 把剥掉最外层的洋葱连同产生的随机数填充组成与原来同样大小的洋葱数据包, 这样在中间任何一个路由器上看到的 information 都不相同, 但包的大小相同, 从而加大了攻击者分析的难度。



图 6.13 洋葱路由技术实现过程

Crowds 是一种基于组群的结构, 它利用 P2P 网络信任分布与均衡负载的特点, 为 Web 浏览用户提供发送者匿名的保护。其基本思想是将发送者隐藏在一个组群中, 通过转发节点的随机转发达到匿名的目的。与洋葱路由机制不同的是, Crowds 机制的下一跳转发节点是由随机转发概率在传输过程中来决定的, 因此其路径随机变化。图 6.14 显示 Crowds 系统中节点进行转发过程的示意图, 其中, J1~J6 是 Crowds 系统成员, 各成员在逻辑上是全连通的, Server 表示为 Web 服务器。当 J1 想要匿名访问 Web 服务器时, 根据转发概率进行转发, 选择转发路径 J1→J2→J4→J5→J5→J4→J6→J5→Server, 从而完成整个访问, 达到匿名的目的。从图还可以看出, 每个成员可以在路径上出现多次, 这只与转发概率有关系, 可以自己选择自己作为下一跳转发节点。图中, 带箭头的直线表示匿名请求转发的方向。另外, WonGoo 协议将分层加密思想和随机转发方式相结合, 任意两点间的通信采用定向与随机选择两种方式选择中间节点, 从而提供发送者匿名、接收者匿名和发送者-接收者的不可链接性。这类匿名通信系统也适用于低延迟网络, 但由于采用随机概率随机转发, 造成转发路径可能非常长, 从而带来系统负载和延迟开销。

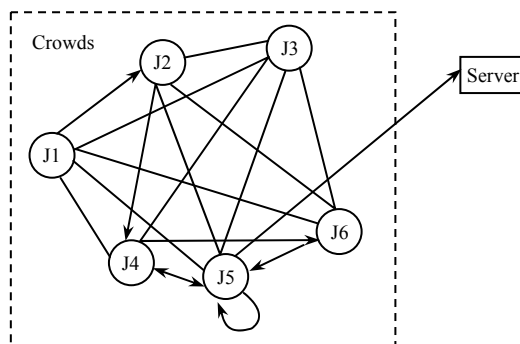


图 6.14 Crowds 系统转发过程示意图

3. 无线移动自组网匿名系统

随着无线移动自组网应用的增加, 无线匿名自组网上的匿名需求也越来越受到研究者的关注。由于移动自组网与有线网络结构的不同, 导致有线网上的匿名通信协议并不适合与移动自组网, 因此, 越来越多的研究者着手研究, 提出了一些典型的匿名协议,

如 ANODR 协议、MASK 协议等。ANODR 协议是一种不可追踪路由的匿名按需路由协议，其基本思想是通过广播找到到达接收者的路径，并通过该路径构建洋葱包进行消息的传递；ASR 协议是 Zhu 等提出的简单有效的匿名安全路由协议；AnonDSR 匿名通信协议利用安全参数建立、匿名路由发现及匿名数据传输等三个过程实现了抗流量分析的安全匿名性；ZAP 协议是基于区域的匿名位置路由协议，它采用了基于组群的思想，将接收者隐藏在一组实体 AZ 内来实现接收者匿名的目的；ODAR 协议是利用布鲁姆过滤器（Bloom Filters）的源路由匿名通信协议，由于路径保存在 Bloom filter 中，造成空间的浪费，以此换取匿名的性能；SDAR 协议同样也是利用广播和洋葱包进行匿名通信的，但其中间转发节点的选择是根据信誉度来确定的，而且洋葱包是接收节点根据 RREQ 包中所携带的中间转发节点信息一次性产生的，并通过返回路径传送到发送者；MASK 协议利用中心可信任权威服务器和匿名邻居身份认证机制，在发送者和接收者之间建立了多条路径，可达到无条件传输关系匿名和节点位置的不可定位性。

6.4.3 匿名性能与效率

匿名性能与效率主要是指匿名系统所能达到的匿名性能强度以及系统为匿名所付出的代价。强匿名性能一直是匿名技术研究所关注的问题，强匿名性是指当攻击者能够进行全局窃听或攻克相当比例的系统成员情况下，系统仍能保持很好的匿名性。一般研究匿名通信系统的强匿名性都是在一定攻击环境下来进行的，即在具有何种能力的对手的攻击下，用户能够保持什么样程度的匿名性。基于 DC-Net 的匿名系统尽管提供了理论上可证明的匿名，但是解决冲突的系统效率很低，而大量的源重写系统在不同的实现中，匿名性能有很大的差异，抗攻击能力有限。研究针对强攻击的匿名技术，特别是针对全局窃听、数据流分析以及主动攻击的匿名技术，是匿名技术研究的一个重要方面。同时，在获得强匿名性的时候，系统往往会为匿名付出一定的代价，如系统延迟、网络带宽的消耗、网络管理的开销等。研究匿名系统性能与系统效率的折衷也越来越受到研究者的关注，力求能在获得强匿名性的同时付出最小的系统代价。

匿名性的获取是各种匿名技术研究的目标，如何衡量一种匿名通信技术或机制的匿名性能一直是研究人员所关注的问题，目前已有多种匿名性度量的方法，最经典的分析匿名度的方法就是匿名集大小，潜在发送者 / 接收者越多，被攻击者猜中的可能性就越小。但是当匿名集中的成员发送的消息很小时，攻击者能很快猜出发送者或接收者，所以此时匿名集大小并不能很好的描述一个匿名系统的匿名性，对匿名性的衡量应该考虑发送消息和接收消息的分布广度，分布越广，匿名性将越好。Reiter 和 Rubin 利用概率对 Crowds 系统匿名度进行了定量和定性度量，在从攻击者角度将发送者的匿名性能分为了 6 个等级后，采用攻击者得到匿名对象可能的概率 p 来衡量匿名程度， $1-p$ 即是同一匿名集中对象的匿名度。Berthold 等人尝试采用 $A=\log_2 N$ 来定性衡量匿名度，其中 N 是系统中的用户数。在这样的定义下，匿名度仅与系统中的用户数有关，没有考虑攻击者可能通过一定的攻击手段获得相关信息后，对匿名集中对象判定的不一致性。因此该匿名度定义不能用来测试系统针对攻击的强壮性。香农的信息论模型提供了一种熵值的匿名度定量衡量方法，熵可以用于表示概率分布的不均匀度，概率分布越均匀，熵值越大，当概率分布等同时，熵值达到最大，攻击者也就最不能确认发送者的真实身份。根据熵的定义，学者们研究了用条件熵和期望熵值衡量匿名度的方法，该方法与前面所提方法不同的是，研究者考虑了攻击者可能观察到的多种不同事件的情况，针对发送者匿

名, 用条件概率 $\Pr\{S=v|F=w\}$ 表示当窃听者收集了信息 w 后, 结点 v 是真正发送者的概率。用熵值的期望 $H^*(S)$ 来衡量匿名度。

目前, 针对于如何获取强匿名性能的问题, 研究者一般采用的方法是先假定攻击者的攻击能力, 然后研究抗强攻击的匿名技术。从攻击者所处的位置来分, 可以分为内部攻击和外部攻击, 内部攻击是从匿名通信系统内部发起的, 攻击者就是组成匿名通信系统的成员或已控制了其中部分组成成员, 如发送者、接收者或中间转发节点等; 外部攻击是在匿名通信系统外部或控制匿名通信系统底层的通信介质来发起攻击, 如外部窃听等。从攻击的行为来分, 可以分为被动攻击和主动攻击, 一个被动攻击者一般只是监听网络流量, 而一个主动攻击者具有任意修改、增删或延迟所传送的信息包等能力。从攻击者所占用的通信资源来分, 可以分为静态攻击和自适应攻击, 静态攻击在攻击过程中所占用的资源不变, 而自适应攻击在攻击过程中可能不断改变占用的资源和策略。从攻击者的攻击能力来分, 可以分为全局攻击和局部攻击, 全局攻击能从整个匿名系统出发, 发现整个匿名系统的数据, 而局部攻击者只能在一定范围内进行攻击。对于匿名通信系统的攻击者来说, 可能同时是全局的、外部的、被动的、静态攻击者, 因此, 必须采取不同的对策。典型的攻击方法包括有流量分析攻击、前驱攻击、合谋攻击、女巫攻击等。流量分析攻击是一种普遍有效的攻击手段, 在通过对网络流量的监听和分析后, 攻击者可从流量数据经过的时间、大小、类型等方面发现发送者和接收者之间的关联性或发现流量数据进出转发节点的关联性, 从而推测出数据的发送者或接收者, 达到破坏匿名通信的目的。时间攻击、报文整形攻击、交集攻击、报文标记攻击、暴力攻击和泛洪攻击等都属于此类攻击手段。前驱攻击是一种针对于匿名通信系统的内部联合攻击, 攻击者利用匿名系统成员加入策略, 加入到系统内, 作为匿名系统的正式成员, 加入到系统匿名信息转发过程中, 从而可以收集并共享信息以推断通信双方的身份, 发现发送者或接收者。前驱攻击者认为, 发送者在多次路径重置后在其前驱位置上出现次数最多, 因此, 只要在通信期间重路由路径重置次数足够大, 根据大数定理, 任何匿名通信协议都将被攻破。合谋攻击是攻击者利用系统漏洞, 加入到系统中, 多个攻击者共同串通, 获取各用户所应享受的更大权限信息或更多的系统信息, 再从这些信息中推断出匿名隐藏关系。对于合谋攻击, 唯一解决的方法就是尽可能地将合谋攻击者挡在匿名通信系统的外面或转发路径之外。女巫攻击是指攻击者假冒成多个正常节点, 并加入到匿名系统, 通过合作获取匿名系统中数据转发的额外信息, 从而推断出发送者或接收者或发送者和接收者之间的路径信息。这种攻击主要是利用匿名通信系统中的冗余信息来达到目的的, 由于移动自组网匿名系统中在建立转发路径时普遍采用广播或组播技术, 因此, 女巫攻击是移动自组网匿名系统的一种重要攻击。除了上述攻击之外, 信息编码攻击、信息量攻击、局部视图攻击等也可以对匿名通信机制造成破坏。另外, 在匿名通信系统和协议的应用中, 还有一些特别地专门针对匿名系统安全弱点的攻击, 例如, Mix 服务器的管理员可能泄露信息, 使管理员无法将 Mix 输入信息与输出信息关联起来。

强匿名性必定带来一定的代价, 在匿名技术最开始的研究中, 往往在强调匿名性的时候忽略了系统效率和扩展性的问题。例如为了获得强匿名性, 普遍采用多重非对称加密技术、重路由技术、填充包和广播技术等。在系统管理上大部分采用了集中管理的技术, 也使得系统效率低、可扩展性差。在近期的研究中, 开始将关注点放在了效率等方面, 力求在获得强匿名性的同时得到高的效率。为此, 许多研究者们开展了匿名系统性能与系统效率的折衷研究。

6.5 本章小结

本章介绍了其他一些主要的信息隐藏技术，包括隐蔽信道、阙下信道、低截获概率通信和匿名通信。

对于安全信息系统来说，机密信息的泄漏将会造成无法挽回的损失。对于隐蔽信道这种危害性极强的信息泄漏途径，必须做到提前预防、尽早发现、及时处理，以保证信息的安全。随着对隐蔽信道研究的逐步深入，隐蔽信道分析技术也逐渐成熟。为了保证系统安全，降低隐蔽信道的危害，未来的系统开发应该更注重设计阶段的形式化描述，从源头上减小隐蔽信道发生的可能性。同时，要规范系统实现过程，严格按照形式化描述，尽最大可能消除代码中引入的潜在隐蔽信道。

阙下信道技术就是依托于公钥密码技术而发展起来的一个新的领域，是秘密中的秘密。根据阙下信道的特点，我们把它归入了信息隐藏技术中。实际上，与其说它是一种信息隐藏技术，不如说是一种密码技术，因为所涉及的技术都是数学的和密码的。随着签名认证等技术的广泛应用，阙下信道技术的应用平台已经形成，有很多学者开始关注这一领域，这是喜人的。然而，阙下信道技术的进一步发展和推向应用还需要一定的时间，寻找阙下信道技术的新的应用对推动该领域的发展非常重要。

随着人们对扩频技术研究的逐步深入，扩频体制各种潜力正被挖掘出来，由于扩频技术采用了扩频编码调制和相关处理技术，具有低的发射功率谱密度，可在低信噪比环境下工作，抗杂波干扰和信号衰落、可多址复用且有保密性，所以扩频技术在通信、雷达、电子对抗、卫星通信各个领域扮演着重要的角色，越来越被人们重视。流星突发通信正处在蓬勃发展之中，目前已有商用产品的出现，其技术性能不断完善和提高，由于高速电子开关技术、先进计算机和自动应答系统的出现，能及时准确地检测到流星余迹，并可在瞬间可靠地转发信息，从而把这一特殊的通信技术推向成熟。据有关专家预测，流星突发通信技术最终有可能代替卫星通信系统，使全球范围内的通信变得更加便宜、可靠和保密。

因特网上的保密通信和数字产品版权保护等方面强烈需求已成为匿名通信技术研究的强大推动力，然而目前报道已有犯罪分子也利用这些技术在因特网上进行秘密交流和招募新人，这就给社会安定埋下很大的隐患。所以匿名通信技术是一把名副其实的双刃剑，匿名通信技术需要人们进一步的研究。



习题

1. 请介绍隐蔽信道的概念和分类。
2. 请介绍阙下信道的概念和构造方法。
3. 请介绍扩频通信的各种工作方式。
4. 请介绍匿名通信的概念和体系结构。

第 7 章

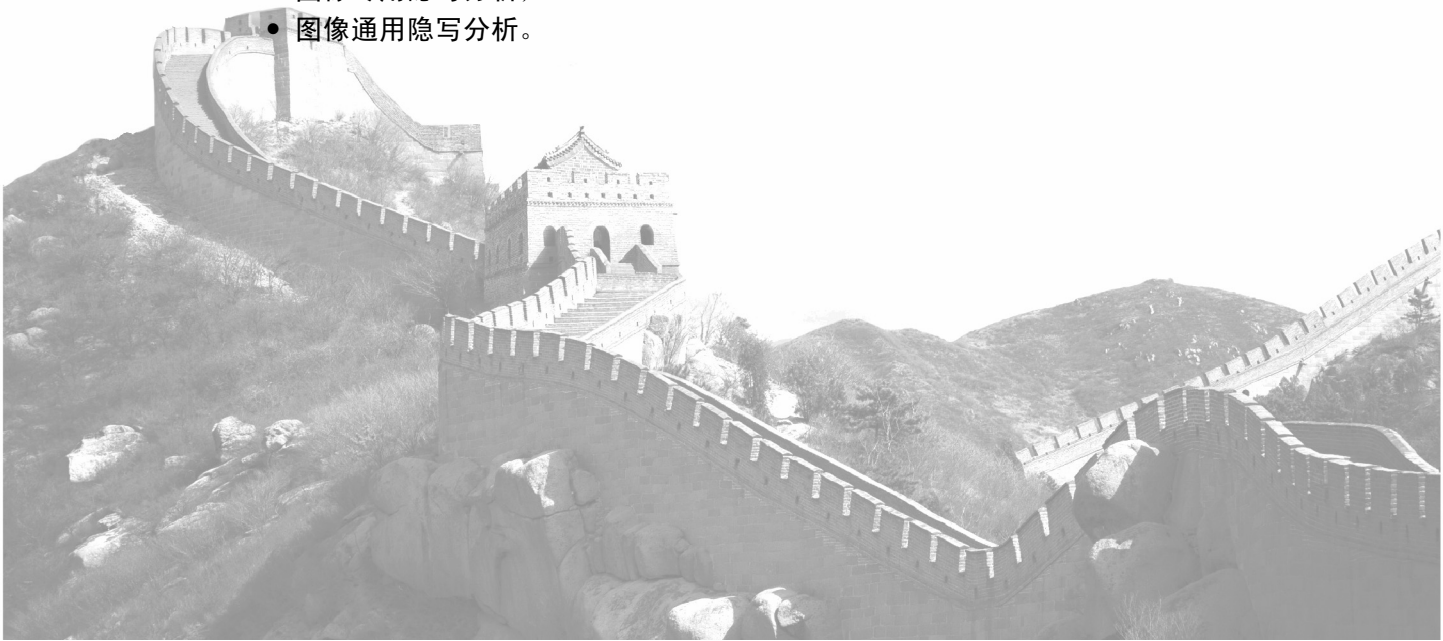
隐写分析技术

本章引言

第 2 章中介绍的隐写术是将秘密信息以一定的方式嵌入到公开的载体对象中，从而隐藏了秘密信息的存在，可以实现保密通信。本章介绍隐写术的对立面—隐写分析，它是对隐写术的检测和攻击，即对可疑的隐写对象进行攻击，以实现秘密信息的检测、破坏，甚至提取出秘密信息。隐写分析技术有利于防止隐写术的非法应用，能够在防止机密资料泄漏，打击恐怖活动和揭露非法信息等方面起到非常重要的作用，有益于保证国家安全和社会稳定。隐写分析不仅具有重要的应用价值，在学术研究上也具有重要的意义。隐写分析可以揭示当前隐写方法存在的缺陷，从而对隐写方法的安全性进行评测。因此，隐写分析方法对于如何设计安全的隐写方法具有指导意义。

本章重点

- 隐写分析的基本概念和分类；
- 隐写分析算法的评价指标；
- 图像专用隐写分析；
- 图像通用隐写分析。



7.1 隐写分析基本概念和分类

7.1.1 基本概念

隐写分析是对隐写术的攻击，目的在于揭示隐写对象中是否存在秘密信息以致破坏保密通信。对于隐写术来说，不可感知性是其首要目标。尽管人的感觉不能发现隐写对象中秘密信息的存在，并不代表计算机不能发现。嵌入的过程或多或少总是会改变原始载体对象，从而破坏载体对象的一些固有特性。而其统计特性可以通过计算机获得，并被用来分析载体中是否被嵌入过信息。隐写分析的主要目标是根据载体对象和隐写对象的某一特性或者统计特性的不同来检测隐写对象中是否有秘密信息，如果判断含有秘密信息则截获，进而预测秘密信息的长度甚至提取秘密信息。该过程通常可以划分成三个阶段：首先，判断隐写对象中是否隐藏有秘密信息；其次，在判断出隐写对象中含有隐秘信息的基础上，使用针对性隐写分析方法判断可能使用的隐写方法，并确定被嵌入秘密信息的长度；最后，确定隐写方法和嵌入密钥来提取秘密信息。总之，隐写分析是利用秘密信息的嵌入可能引发载体数据分布特性或统计特性的改变，分析在信道中获得的可疑隐写对象，从而检测、估计并提取出隐藏的秘密信息，其框架如图 7.1 所示。

由于隐写分析者只能得到可疑隐写对象，而提取秘密信息所需的载体对象、隐写算法、密钥等都是未知的。因此目前对隐写分析的研究主要集中在第一和第二阶段，即大都针对确定可疑隐写对象中是否存在秘密信息或确定秘密信息的长度和位置。在图 7.1 中，前一阶段检测秘密信息是否存在属于被动隐写分析，该阶段的研究成果最为显著。因为从某种意义上说，当秘密通信的事实被发现后，隐写分析者可通过破坏秘密通信信道以阻止该过程。主动隐写分析是指估计嵌入的秘密信息的长度、嵌入的位置及嵌入算法中使用的密钥和相关参数，最终提取秘密信息。该阶段的研究目前尚处于发展阶段。

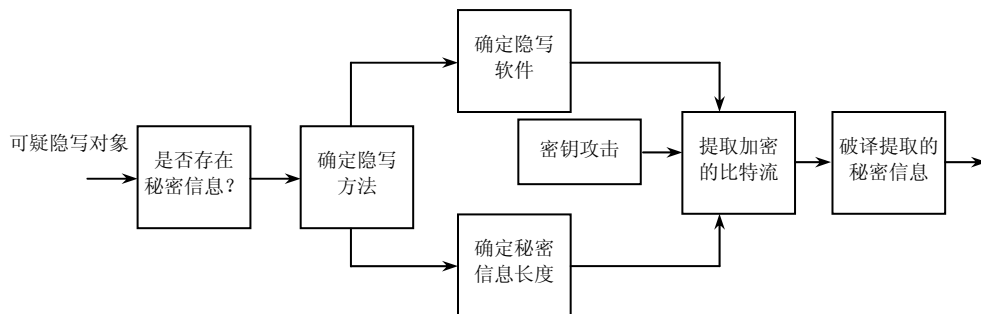


图 7.1 隐写分析系统框图

7.1.2 分类

根据隐写分析的适用性，隐写分析技术可分为专用隐写分析技术和通用隐写分析技术。专用隐写分析方法是针对某种具体的嵌入方法进行的，提取出其专有特征，并根据这些专有特征来判别是否隐写。如经典的 χ^2 分析法和 RS 分析法是针对 LSB 替换的专用隐写分析方法，不仅可以检测出待测可疑隐写对象中是否含有秘密信息，还能估计出秘密信息的嵌入量。通用隐写分析方法是根据通用特征进行判别，这些通用特征是独立于具体的隐写算法之外的。例如基于音频质量测度的隐写分析方法，它是从隐写造成的音

频失真的角度来考虑的,并没有考虑到具体的隐写方法。专用隐写分析技术可准确检测出某种特定的隐写方法,但适用性较低。通用隐写分析技术也许对于某一种隐写方法的检测准确性不如专用隐写分析技术,但其适用性较高。

1. 根据隐写分析达到的效果或目标分类

根据隐写分析达到的效果或目标分类,可把隐写分析分为以下三类。

(1) 被动攻击(也叫检测技术)

目的是检测可疑隐写对象中是否存在秘密信息。检测技术是隐写分析技术的第一步,也是现阶段隐写分析技术研究的主要内容。

(2) 主动攻击(也叫攻击技术)

目的是对隐写对象进行破坏,使信息无法被提取。这种方法实现简单,对于破坏保密通信非常有效。在隐写技术中,对隐写对象的几何变形、压缩、加噪等操作都可能会破坏隐藏的信息,而造成无法提取。

(3) 提取攻击(也叫破解攻击)

目的是获得秘密信息的内容,即截取隐写对象后,分析出隐藏的秘密信息。破解技术与秘密信息的嵌入方式有很大的关系,涉及数据加密、信号处理等相关知识,因此具有较大的难度。

实际上,后两者可以统称主动隐写分析,主动隐写分析比被动隐写分析的要求更高、难度更大。在上面这3个层次中,破解攻击的难度最大,现阶段还没有进行深入研究,目前研究的重点主要集中在检测和攻击技术上。检测技术又可分为对比检测技术和盲检测技术。对比检测技术简单、可靠,但通常由于原始载体无法获取,实际应用的意义不大。盲检测技术由于不需要原始载体而具有更加广泛的应用前景。

2. 根据隐写分析者能够获得的信息

根据隐写分析者能够获得的信息,也可将隐写分析技术分为以下六类。

(1) 唯隐写对象分析

分析者只能得到可能嵌入秘密信息的隐写对象,并不知道可能采用的隐写方法与隐写内容。这种隐写分析方式难度较大,主要任务是检测秘密信息是否存在。

(2) 已知载体对象分析

分析者既能获得隐写对象,同时还拥有原始载体对象。这种类型的分析比较简单,只需检测载体对象与待测的可疑隐写对象之间是否存在差异,从而判断出待测可疑隐写对象是否含有秘密信息,如果能够破解密钥,还能够提取出秘密信息。但是在一般情况下,分析者是无法得到载体对象的,因此这种隐写分析方式应用的场合不多。

(3) 已知秘密消息分析

在某一点上,隐藏的秘密信息可能为分析者所知。但仅通过秘密信息来判断可疑隐写对象是否含有该秘密信息是非常困难的,其难度甚至等同于唯隐写对象攻击。该隐写分析方式通常为其他未知秘密信息时的隐写分析提供一些参考。

(4) 选择秘密信息攻击

分析者可以选择一个秘密信息,并利用某个隐写算法或工具对该秘密信息产生隐写对象,然后根据不同的隐写算法对隐写对象所产生的特征进行分析,以判断出隐写对象中所使用的隐写算法或工具。

(5) 选择隐写对象攻击

分析者能够到隐写对象,并且知道所采用的隐写方法或隐写工具。选择一些特定的

载体对象，并应用这种隐写方法或工具来产生隐写对象，通过对这些隐写对象的特征进行分析来判断待测可疑隐写对象中是否存在秘密信息。

(6) 已知隐写攻击

秘密信息、载体对象和隐写算法都已知，在这种方式下进行隐写分析是比较容易的。

显然，上述第一种情况在技术上最具挑战性，是隐写分析的重要研究内容。不妨说，成功地实现针对任何对象、任何隐写方法的盲分析是分析者要达到的最终目标。然而对隐写算法和隐藏内容一无所知的全盲分析往往非常困难，因此，迄今为止人们常对一些有效的隐写方法和特定的对象研究采取具有针对性的分析技术。

3. 根据分析过程中所采用的方法

另外，根据分析过程中所采用的方法还可以将隐写分析分为以下三种。

(1) 感官检测

为了能够抵抗攻击，一般在载体对象的比较敏感的区域隐藏秘密信息，但同时也可能产生感官痕迹，从而暴露秘密信息。人类的感官具有感知和分辨噪声的能力，这种方法首先对待测可疑隐写对象进行一定的预处理，然后直接利用人类感官来进行判断。本方法的可靠性差，只能针对简单的隐写方法，与具体的载体有很大的关系，并且需要人工判别，不能自动检测。随着隐写技术的发展，隐写引入的噪声越来越难被人类感官所感知，因此目前一般不采用该方法。在数字载体的失真和噪声中，人类可感知的失真或模式最易被检测到。辨别这种模式的一个方法是比较载体对象和隐写对象，注意可见的差异。如果没有载体对象，这种噪声就会作为载体的一个有机部分而不被注意。感官检测的思想是移去载体信息部分，这使人的感官就能区分剩余部分是否有嵌入秘密信息或仍然是载体的内容。当然，因为人的感知有一定的冗余度，且隐写算法的首要任务就是不能超出人类视 / 听觉冗余度，人类感官系统不易察觉到秘密信息的存在，但这种变形和降质确实存在，可以配合对载体的处理，使得感官检测能达到一定的成效。感官检测不适合计算机的自动化分析检测，尤其是当分析的媒体来源于网络时，要求设计的分析算法必须满足实时性和低漏警率的要求。

(2) 统计检测

这种分析方法首先要能够得到原始载体的理论期望分布，然后与待测可疑隐写对象的样本分布做比较，根据它们之间是否存在差别来检测隐写。隐写改变载体数据流的冗余部分虽然不改变感觉效果，但是却经常改变原始载体数据的统计性质。通过判定给定可疑隐写对象的统计性质是否属于反常情况，从而可以判断是否含有秘密信息。统计分析方法的前提条件是能够得到原始载体对象的理论期望分布，但在许多情况下，这种分布难以得到。因为基于不同格式载体的隐写方法多种多样，所以对它们进行统计攻击的具体方法不同。

(3) 特征分析

特征分析就是对载体对象与隐写对象分别提取相应的特征（标志特征、格式特征或统计特征），找出两者在特征上的差异，从而对载体对象与隐写对象进行区分。这种特征可以是感官的、统计的或可以度量的。这种方法是目前最为常用的隐写分析方法。广义地说，进行分析所依赖的就是特征，这种特征必须根据具体的应用情况通过分析发现，进而利用这些特征进行分析。感官上的、格式上的特征一般来说较明显，也较容易，如基于文件格式中空余空间的隐写分析，磁盘上未使用区域的隐写分析等。其他较复杂的隐藏特征则要根据隐写算法进行数学推理分析，确定载体对象和隐写对象的度量特征差异。通过度量特征差异分析隐写往往还需借助对特征度量的统计分析。

7.2 隐写分析算法的评价指标

在隐写分析算法的应用中,面对不同问题,其侧重点可能也不尽相同。例如在网络环境下进行大流量数据检测时,为有效节省计算资源,算法对虚警率要求相对较高:即宁愿牺牲一定的漏报率,也要保证把虚警率控制在某个水平以下。本节从如何有效应用隐写分析算法为出发点,介绍隐写分析算法的几种量化评估指标,包括准确性、可靠性、适用性、分类代价、实用性和计算复杂度。

7.2.1 可靠性和准确性

前面已经提到,按照攻击效果,隐写分析算法一般分成两大类:被动分析算法只能判断可疑隐写对象是否含有秘密信息,但不能估计秘密信息的长度;主动分析算法不仅可以判断可疑隐写对象是否含有秘密信息,还可以估计秘密信息的长度,也称之为定量隐写分析算法。到目前为止,国内外学者提出了较多的定量隐写分析算法。定量隐写分析算法可以估计嵌入秘密信息长度,从而可以计算出秘密信息的嵌入率。这里,嵌入率是指嵌入的秘密信息占载体信息容量的比率。这些算法实际上都是构造一个估计嵌入率的统计量,该统计量是一个需要评价的指标,可以分别从无偏性和有效性来评价。以算法实验得到的嵌入率估计值和真实值的差为评价对象,可以引入如下两个指标。

① 可靠性:表征隐写分析算法对嵌入率估计值的离散程度或平稳程度,用于度量统计量的有效性。通常,采用标准差来衡量统计量的可靠性。② 准确性:表征算法对嵌入率估计值的准确程度,用于度量统计量的无偏性。通常,采用均值来衡量统计量的准确性。

可靠性表征了定量隐写分析算法检测不同可疑隐写对象时的稳定程度,准确性表征了算法检测不同可疑隐写对象时的准确程度。当待检测对象(如纹理图像)比较复杂时,算法的可靠性是首先需要考虑的问题,因为只有可靠性高的算法才能保证检测结果具有较高的可信度。另一方面,在某些情况下,例如在研究某类隐写术的提取工作过程中,针对待检测对象,提取算法需要得到较准确的嵌入率大小,从而降低提取工作的计算复杂度,此时准确性就是需要着重考虑的指标。

7.2.2 适用性

1. ROC 分析简介

隐写分析算法的检测过程其实就是一个二值分类过程:经过算法的检测,被检测的可疑隐写对象被分成两类,载体对象(即不含秘密信息的检测对象)或者隐写对象(即含有一定秘密信息的检测对象)。虚警率是把载体对象错误判断为隐写对象的概率,表示为 $P(\text{隐写对象}|\text{载体对象})$,常用 α 表示,漏报率为把隐写对象错误判断为载体对象的概率,表示为 $P(\text{载体对象}|\text{隐写对象})$,常用 β 表示,检测率是把隐写对象正确判断为隐写对象的概率,表示为 $P(\text{隐写对象}|\text{隐写对象})$,常用 $1-\beta$ 表示,参见图 7.2。

在二值分类问题中,ROC 分析是比较常用的分析方法。描绘虚警率和检测率两者关系的曲线称为检测器接收操作特性(Receiver Operating Characteristic, ROC)曲线。在 ROC 分析中,全面衡量分类效果的一个量称为全局检测率,定义为

$$P_t = 1 - P_e \quad (7.1)$$

其中, P_e 为平均错误概率,其定义如下

$$P_e = \alpha \cdot P(\text{载体对象}) + \beta \cdot P(\text{隐写对象}) \quad (7.2)$$

其中, $P(\text{载体对象})$ 和 $P(\text{隐写对象})$ 分别表示从被检测对象空间中取到载体对象和隐写对象的概率, α 表示虚警率, β 表示漏报率。在 ROC 分析中, 全局检测率 P_t 的大小用 ROC 曲线图与横坐标包含区域的面积 AUC (Area Under Curve) 表示。虚警率与检测率相等时所对应的 ROC 曲线如图 7.3 所示, 其中 AUC 为 45 度对角线与横坐标围成区域的大小, 此时对应的全局检测率为 0.5, 属于随机猜测: 判断正确和错误的概率均为 0.5, 此时的检测器无效。当全局检测率即 AUC 达到 0.85 及以上时, 认为检测器性能良好。

被检测对象的分类结果	被检测对象的真实类型	
	载体对象	隐写对象
载体对象	载体对象正确判断为载体对象	隐写对象错误判断为载体对象
隐写对象	载体对象错误判断为隐写对象	隐写对象正确判断为隐写对象

图 7.2 虚警率、检测率和漏报率的示意图

2. 适用性

隐写分析算法的一个重要目的是检测可疑隐写对象是否隐藏有秘密信息, 而可疑隐写对象形式多样, 一般由载体对象、隐写算法和秘密信息等因素确定。为了考察隐写分析算法对不同待检测对象的适用程度, 引入适用性指标, 它用能够达到有效检测隐写对象的种类来表示, 隐写分析算法可以有效检测的隐写对象种类越多, 其适用性越好。

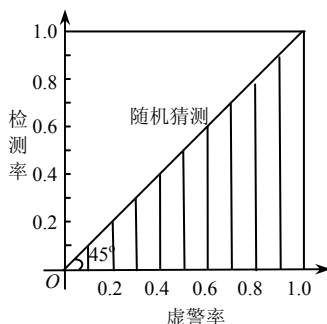


图 7.3 随机猜测对应的 ROC 平面

隐写分析算法对某类对象的检测效果表征着对应的适用性程度, 检测效果则利用上面提到的全局检测率 AUC 量化表示, 规定有效检测为 AUC 大于等于 0.85。

隐写对象一般由以下三个因素共同决定而成。

(1) 隐写算法

隐写算法指嵌入秘密信息时所采用的规则, 包含可能需要使用的隐写密钥。常见隐写算法有: 随机 LSB 替换、连续 LSB 替换、随机 LSB 匹配、BPCS (比特平面复杂度分割)、SSIS (Spread Spectrum Image Steganography)、随机调制隐写等。

(2) 载体对象

载体对象就是用于嵌入秘密信息的可公开的数字媒体。常见的载体对象有图像、音频、文本和视频等。图像由于应用的广泛性, 已经成为最常用的载体对象之一。载体对象体的统计特征往往呈现出多样性, 比如 BMP 灰度图像, 可以按纹理复杂程度分类, 也

可以按图像的大小来分类。因此,确定载体对象不但需要考虑载体的数字类型,而且还要考虑对某类数字对象进行的不同分类。

(3) 秘密信息

秘密信息的主要特征是其分布和数据长度。一般情况下,隐写时采用的秘密信息服从一定的概率分布,如两点分布、高斯分布等。经过加密后得到的密文序列为两点分布,其中的0和1近似均衡,取0的概率为0.5,而未加密的明文序列亦为两点分布,其中的0、1不都均衡,取0的概率不都为0.5。秘密信息嵌入的多少可以用秘密信息的长度占载体对象的比例即嵌入率来表示,嵌入率是影响算法检测效果的重要因素之一。一般嵌入率越小,载体的修改就越少,隐写分析算法的检测效果越差。定义隐写分析算法能检测的最小嵌入率为算法达到有效检测时对应隐写对象的嵌入率大小。

通过上面三个因素的两两组合,可以得到不同类型的被检测对象。针对不同的被检测对象,利用隐写分析算法的检测结果进行ROC分析,并计算出全局检测率AUC,AUC的数值就量化表征了算法对此类隐写对象的适用性程度,若达到有效检测,即若AUC大于等于0.85,则算法适用于检测此类隐写对象。隐写分析算法的适用性就用达到有效检测的隐写对象数量来表示。

3. 最小嵌入率的计算

隐写分析算法能检测的最小嵌入率表征了隐写分析算法能有效检测的嵌入率范围,因而是在评价算法的适用性时必须考虑的一个问题。我们固定其他影响检测效果的因素,只考察嵌入率的变化:从某个合适值开始依次减小嵌入率的大小,对应的算法检测效果也递减,分析两者的关系,从而计算出最小嵌入率的大小。根据算法的检测性能,嵌入率在选取的合适范围内变动,如嵌入率变化区间为 $[0, 0.1]$,这个嵌入率区间被分成 n 等分,其中 n 的选取根据所需精度确定,实验表明,一般取 $n=10$ 可以得到较准确的估计。依次选择 n 个嵌入率,利用算法的检测结果分别进行ROC分析,得到 n 个对应嵌入率的AUC,这样得到了一组嵌入率与AUC的数据。实验发现,AUC与嵌入率的关系可以用多项式函数较好的拟合,根据拟合数据,对应AUC等于0.85时的嵌入率,就是隐写分析算法能检测的最小嵌入率。

7.2.3 分类代价

根据上一小节的分析,全局检测率AUC衡量着隐写分析算法的分类效果,但是全局检测率的计算公式并没有涉及虚警率和漏报率的权值大小,此时假设的是虚警率和漏报率造成的代价是相同的。其中,虚警率的代价是指资源花费,漏报率的代价是指信息损失。例如,在网络环境下的一个实时检测系统,其每天处理可疑对象的数据量很大。一方面,若算法的虚警率较高,将产生大量载体对象需要进一步处理,极大的占用存储空间和计算资源,此时虚警率造成的代价就是这些无意义的资源花费。另一方面,若算法的漏报率较高,将会遗漏掉很多隐写对象,可能造成重要价值的信息损失。根据实际应用对资源花费和信息损失的不同要求,全局检测率已经无法准确判断算法在实际情况下的分类效果。基于此,可引入分类代价分析,目的是研究在不同的应用需求下,隐写分析算法造成的实际代价。根据实际情况对虚警率与漏报率的不同要求,分类代价分析可得到隐写分析算法对应的代价变化情况,从而选择代价最小的算法进行检测。代价最小的算法对应的资源花费与信息损失会达到合适的折衷。实际应用对资源花费和信息损失的不同要求可以转化为虚警率和漏报率的不同要求:实际应用需要控制虚警率在 α_0 以内,控制漏报率在 β_0 以内,则虚警率与漏报率的代价之比为 β_0 / α_0 。ROC分析的另一种

表现形式称为花费分析（COST 分析），由此我们定义以下两个概念。

1. 概率花费函数（Probability Cost Function, PCF）

$$\text{PCF}(+) = \frac{p(+)\cdot C(-|+)}{p(+)\cdot C(-|+) + p(-)\cdot C(+|-)} \quad (7.3)$$

其中 $C(+|-)$ 和 $C(-|+)$ 分别是虚警率和漏报率的代价， $p(+)$ 和 $p(-)$ 分别是取到隐写对象和载体对象的概率。类似的可以定义 $\text{PCF}(-)$ 如下

$$\text{PCF}(-) = \frac{p(-)\cdot C(+|-)}{p(+)\cdot C(-|+) + p(-)\cdot C(+|-)} \quad (7.4)$$

显然， $\text{PCF}(+) + \text{PCF}(-) = 1$ 。若设虚警率和漏报率的代价之比为

$$\eta = \frac{C(+|-)}{C(-|+)} \quad (7.5)$$

则 PCF 的定义可以转化为如下形式

$$\begin{cases} \text{PCF}(+) = \frac{p(+)}{p(+) + \eta \cdot p(-)} \\ \text{PCF}(-) = \frac{\eta \cdot p(-)}{p(+) + \eta \cdot p(-)} \end{cases} \quad (7.6)$$

2. 正规化期望代价（Normalized Expected Cost, NEC）

$$\text{NEC} = (1 - P_T) \cdot \text{PCF}(+) + P_F \cdot \text{PCF}(-) \quad (7.7)$$

其中 P_T 和 P_F 分别为 ROC 曲线的一个点对应的检测率和虚警率。因为 $\text{PCF}(+) + \text{PCF}(-) = 1$ ，故 NEC 可以转化为如下形式

$$\text{NEC} = (1 - P_T - P_F) \cdot \text{PCF}(+) + P_F \quad (7.8)$$

显然可以得到

$$\text{NEC} = \begin{cases} P_F & \text{当 } \text{PCF}(+) = 0 \\ (1 - P_T) & \text{当 } \text{PCF}(+) = 1 \end{cases} \quad (7.9)$$

即概率花费函数为 0 时，正规化期望代价为虚警率；概率花费函数为 1 时正规化期望代价为漏报率。

以 $\text{PCF}(+)$ 为横坐标，以 NEC 为纵坐标作图得到的就是花费分析对应的 COST 曲线图。该曲线反映了随着虚警率和漏报率的代价比 η 的变化，其正规化期望代价 NEC 的变化趋势。而实际情况往往固定 η ，考察虚警率/漏报率和正规化期望代价的变化关系，将 COST 曲线图的横坐标转化为虚警率/漏报率，这样得到的曲线图称为分类代价曲线图。以上的分析称为分类代价分析，具体过程为：根据实际情况对虚警率和漏报率的不同需求，设两者分别被控制在 α_0 和 β_0 以内，确定两者的代价之比 $\eta = \beta_0 / \alpha_0$ ，再根据载体对象和隐写对象的比例确定 $p(+)$ 和 $p(-)$ ，根据式 (7.6) 和式 (7.8)，将 ROC 曲线图转化成分类代价曲线图。分类代价曲线图可以确定：当虚警率/漏报率属于任意确定的区间时，隐写分析算法的分类代价，以及在虚警率/漏报率的具体要求下，不同算法分类代价的优劣程度。据此可以选择最合适的算法以满足实际应用对虚警率和漏报率的不同需求。

7.2.4 实用性和计算复杂度

实用性指的是隐写分析算法能够在实际中使用的能力，包括现实条件是否支持，检

测过程是否需要人工干预,检测的实时性程度以及检测结果是否稳定等方面。

隐写分析算法的计算复杂度是一个重要的评价指标,一般分为时间复杂度和空间复杂度,分别影响着算法的运行速度和存储空间。例如在网络环境下,处理的大流量数据需要算法能够快速检测,计算复杂度高的算法会严重降低资源的利用效率,甚至导致算法的应用失败。我们主要考察隐写分析算法的计算复杂度,计算复杂度分析可以评估算法对资源的利用效率,为应用算法提供了参考。另外,通过分析影响计算复杂度的因素,我们提出了样本量分析,它可作为一种降低计算复杂度的方法。

1. 时间复杂度

隐写分析算法的时间复杂度指算法在处理数据过程中所需要的时间。可通过分析算法的实现步骤从而计算得到,以计算机中的最小处理单元——位操作数目为单位。

2. 空间复杂度

隐写分析算法的空间复杂度指算法在处理数据过程中所占用的存储空间,即所需的内存大小。理论空间复杂度是通过分析算法实现步骤计算得到,以字节(Byte)为单位。

3. 样本量分析

在研究计算复杂度的过程中,我们发现:当隐写分析算法确定后,其计算复杂度的高低主要与待检测对象的大小有关。以数字图像为例说明,隐写分析算法的计算复杂度随着图像的尺寸的减小而降低。既然检测对象的大小决定了计算复杂度的高低,为了降低隐写分析算法的计算复杂度,我们可以取部分图像数据代替整幅图像,作为隐写分析算法的输入。这样,隐写分析算法的计算复杂度随之降低的同时会导致检测效果的降低,这是由于输入到隐写分析算法中的样本量的减小。

为了找到计算复杂度和检测效果两者之间的折衷点,我们考察样本量与检测效果的关系,进行样本量分析。在保证检测效果的情况下,样本量分析可以尽可能的降低用于检测的数据样本量,从而导致计算复杂度的降低。定义隐写分析算法能检测的最小样本量为当隐写分析算法达到有效检测时,隐写对象的样本量大小。通常,对隐写分析算法实验采用不同尺寸的图像,其 ROC 分析得到不同的检测效果,即相同类型的图像,其尺寸越小,检测效果越差。

根据上述分析,样本量分析是为了得到检测效果和计算复杂度之间的折衷而提出的。与适用性分析中的规定一样,算法的全局检测率 AUC 达到 0.85 即为有效检测,针对某个隐写分析算法,其对应的最小样本量指算法达到有效检测时所需要的数据样本量。当然,在实际应用时,可以根据实际对检测效果的需求重新定义有效检测对应的 AUC 值。最小样本量按如下方法计算:将样本量分成 n 等分,其中 n 根据所需结果的精度确定,一般取 $n=10$ 可以较准确的结果,类似前面最小嵌入率的计算过程,依次累加样本量,得到 n 组样本量和全局检测率 AUC 的数据。实验发现,样本量和 AUC 的关系可以用多项式函数较好地拟合,根据拟合数据计算出对应 AUC 为 0.85 的样本量,这个样本量的值就是隐写分析算法能检测的最小样本量。

7.3 图像专用隐写分析

7.3.1 引言

隐写分析的通用基本原理是:根据载体信息在隐写前后统计特性的变化,来分辨出

载体信息是否含有秘密信息。隐写分析算法设计很大程度上依赖于隐写算法的设计，因此根据隐写算法也可以将隐写分析算法进行分类。根据特征的选取与嵌入算法的关系，目前大部分隐写分析算法被归为两大类，即专用隐写分析方法及通用隐写分析方法。前者主要根据隐写图像特征的改变，来提取专用特征进行检测，检测率较高，但实用性不强，只对特定的隐写术有效；后者主要是寻找独立于嵌入算法的统计特征向量，根据载体统计特性的变化判断是否含有秘密信息，它对一系列的隐写算法都有效，实用性较高，但整体检测率较弱。显然，后者是分析者努力达到的目标，但是目前为止，更多分析算法还是主要基于某种嵌入算法的分析，其性能比较优越。

专用隐写分析技术根据具体隐写算法的不同，又可分为空域隐写分析和变换域隐写分析两大类。空域隐写术主要是直接改变图像的像素值来达到隐写的目的。这种方法通常可以在不影响原始载体图像视觉效果的情况下达到高容量的隐写。例如，应用最普遍的 LSB 方法就是将秘密信息直接替换图像像素的最不重要比特位，以降低对图像品质的影响。在提取时则根据嵌入时所使用的方法，从隐写图像中还原出秘密信息。这类算法鲁棒性不高，可通过对统计特性等的检测来分析此类算法。变换域隐写术首先将图像像素值变换成变换域中的系数值，然后将秘密信息嵌入到所选定的系数值上。变换域隐写术较空域隐写术涉及的技术要复杂得多，其隐写方法也是多种多样，从近几年来发表的相关文献中可以看出。其复杂性体现在：变换域选择可以包括 DCT 域、DFT 域、DWT 域等；嵌入算法也各有差别，有的算法修改低频系数进行隐写，有的算法修改中频系数进行隐写，也有的算法利用临近变换域系数之间的关系来进行隐写，还有利用原图像的变换系数之间的关系来进行隐写的。因此针对变换域的隐写分析技术也比较难，但针对一些具体隐写算法，仍可能研究出适当的隐写分析方法。

下面，我们根据对应的隐写术类型，归纳了一些影响较大的典型隐写分析算法。其中，有些空域隐写分析方法也适用于变换域，而有些并不一定适用。

7.3.2 针对空域 LSB 替换的专用隐写分析算法

目前，针对 LSB 替换隐写算法的隐写分析研究发展最为成熟。由于秘密消息中“0”和“1”出现的概率通常相等，而 LSB 隐写算法仅可能将系数值 $2i$, $i \in \{0, 1, \dots, 127\}$ 修改为 $2i+1$ 、或将系数值 $2i+1$ 修改为 $2i$ ，因此系数值 $2i$ 与 $2i+1$ 出现的概率也随数据嵌入率增大而趋近于相等，产生“值对”现象。绝大多数针对 LSB 替换隐写算法的专用隐写分析算法都是基于这个原理而设计的，下面介绍其中三种最典型的分析方法。

1. χ^2 检测方法

卡方 (χ^2) 检测是 Westfeld 首先提出来的^[96]，它通过检验偶值系数的分布是否与奇值系数与偶值系数的算术平均值的分布相同来判断图像数据是否进行了 LSB 替换嵌入。当图像并不是满容量嵌入时，该方法仅能检测嵌入位置是顺序排列的情况。其基本原理可以描述如下：若将 LSB 替换隐写术用于 256 灰度图像，则要么像素不改变，要么像素值从 $2i$ 变为 $2i+1$ ，要么像素值从 $2i+1$ 变为 $2i$ ，而从不进行 $2i$ 到 $2i-1$ 的转化和 $2i-1$ 到 $2i$ 的转化。若定义待检测 256 灰度图像中灰度值 i ($i \in \{0, 1, \dots, 255\}$) 的个数为 N_i 。对于自然图像来讲，如果嵌入了大量的隐密信息，像素值为奇数和偶数出现的频率在理论上应该是一样的，因此有 N_{2i} 与 N_{2i+1} 较为接近，而没有嵌入信息则相距较远。以 512×512 的标准图像 Crowd 为载体 (图 7.4 (a))，嵌入隐秘信息，图 7.4 (b) 和图 7.4 (c) 就是 Crowd 嵌入信息前后的灰度直方图的一部分。可以看出，当嵌入秘密信息后，灰度值对

趋于均匀化, χ^2 检测方法就是利用这种统计特性差异进行隐写分析的。

χ^2 检测方法可以描述如下: LSB 替换隐写不会改变 $N_{2i} + N_{2i+1}$ 的值, 因为灰度只是在 $2i$ 和 $2i+1$ 之间转换。令 $N'_{2i} = (N_{2i} + N_{2i+1})/2$, 构造

$$r = \sum_{i=1}^k \frac{(N_{2i} - N'_{2i})^2}{N'_{2i}} \quad (7.10)$$

r 服从自由度为 $k-1$ 的 χ^2 分布, 其中 k 是 N_{2i} 与 N_{2i+1} 所组成数字对的数量 ($N'_{2i}=0$ 的情况不计算在内), r 越小表示可疑图像含有秘密信息的可能性越大。结合 χ^2 分布的密度函数, 可计算图像被隐写的概率

$$P = 1 - \frac{1}{2^{\frac{k-1}{2}} \Gamma(\frac{k-1}{2})} \int_0^r \left[\exp\left(-\frac{t}{2}\right) t^{\frac{k-1}{2}-1} \right] dt \quad (7.11)$$

当 P 接近 1, 说明可疑图像含有秘密信息。当 P 接近 0, 说明可疑图像不含有秘密信息。

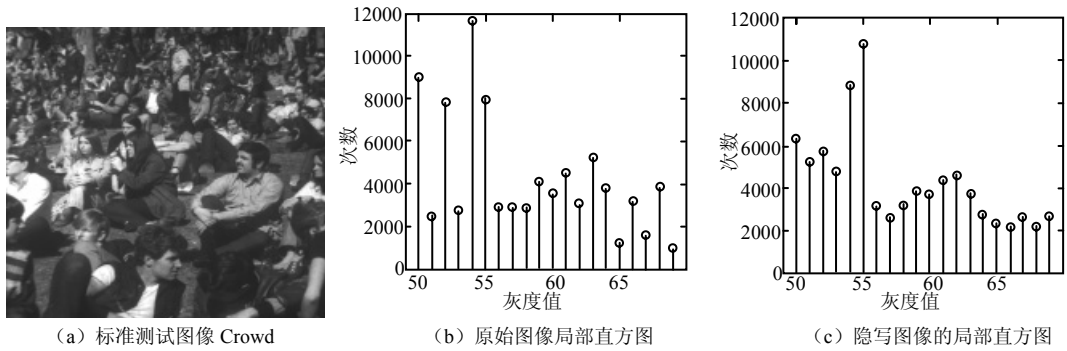


图 7.4 Crowd 图像嵌入秘密信息前后的灰度直方图

2. RS 检测方法

RS (Regular Groups and Singular Groups) 检测法是由 Fridrich 等人提出的^[97], 不仅能对嵌入位置随机排列的情况有效, 而且可以估计嵌入容量。实际上, 在载体图像的 LSB 位平面中, 看似噪声的 LSB 相互之间存在着非线性的关联。而空域 LSB 替换隐写技术破坏了这种关联, 因此这种关联性可用来检测图像中的秘密信息。

假定有一幅 256 灰度载体图像, 将图像中相邻像素分成大小相等的 n 元组

$$G = (x_1, x_2, \dots, x_n) \quad (7.12)$$

其中像素的索引值 $x_i \in W$, 其中 $W = \{0, 1, \dots, 255\}$ 。定义实值函数 f

$$f(x_1, x_2, \dots, x_n) = \sum_{i=1}^{n-1} |x_{i+1} - x_i| \quad (7.13)$$

用这个函数测量像素分组 G 的平滑度和规律性, 像素组 G 噪声越大, f 值越高。利用 LSB 替换隐写方法嵌入信息后, 相当于在原始数据上叠加噪声, 因此嵌入信息后函数 $f(x_1, x_2, \dots, x_n)$ 的值会增加。如果函数 $F(x)$ 满足 $F(F(x)) = x$ ($\forall x \in R$), 则称函数 $F(x)$ 为二轮置换函数。定义如下三个二轮置换函数

$$\begin{aligned} F_1: 0 &\leftrightarrow 1, 2 \leftrightarrow 3, \dots, 253 \leftrightarrow 254, 254 \leftrightarrow 255 \\ F_{-1}: -1 &\leftrightarrow 0, 1 \leftrightarrow 2, \dots, 254 \leftrightarrow 255, 255 \leftrightarrow 256 \\ F_0(x) &= x \end{aligned} \quad (7.14)$$

那么对于每一个 x , 满足

$$F_{-1}(x) = F_1(x+1) - 1 \quad (7.15)$$

对应每个分组 G , 定义如下置换

$$F(G) = \{F_{M(1)}(x_1), F_{M(2)}(x_2), \dots, F_{M(n)}(x_n)\} \quad (7.16)$$

其中, $\{M(1), M(2), \dots, M(n)\}$ 是 n 元组对应的待嵌入秘密信息序列, 其元素值可以为 -1 、 0 或 1 。对分组后的每个像素应用置换函数 F , 相当于在分组中的各个像素中叠加了一部分噪声, 这样就引起了图像像素相关性的变化。根据像素组应用置换函数前后的变化规律, 利用公式 (7.13) 将像素组分成规则组、异常组和无用组三类

$$\text{规则组 } G \in R \Leftrightarrow f(F(G)) > f(G);$$

$$\text{异常组 } G \in S \Leftrightarrow f(F(G)) < f(G);$$

$$\text{无用组 } G \in U \Leftrightarrow f(F(G)) = f(G).$$

首先对检测图像每个分组进行非负置换, 即任意掩码 $M \in \{0, 1\}$, 利用式 (7.13) 分别计算规则组、异常组的个数, 记规则组占全部分组的比例为记作 R_M , 异常组占全部分组的比例 S_M 。然后, 对每个分组进行非正置换, 即掩码 $M \in \{0, -1\}$, 同样的, 分别求出规则组、异常组占全部分组的比例, 分别记作 R_{-M} 和 S_{-M} 。根据公式 (7.15), 如果图像中没嵌入信息, 则有

$$R_M \approx R_{-M} \quad (7.17)$$

$$S_M \approx S_{-M} \quad (7.18)$$

且满足 $R_M > S_M$ 。当嵌入信息后, 大量的实验数据表明式 (7.17) 和式 (7.18) 将不再成立, 但方程不成立的原因至今没有给出理论上的证明, 且随着秘密信息嵌入长度的增加, R_{-M} 、 S_M 随之增加, 而 R_M 、 S_{-M} 却随之减小。即使得 R_M 、 S_M 逐渐靠近, 而 R_{-M} 、 S_{-M} 的差距却随着嵌入 z 长度的增加而增大。当嵌入率达 100% (最低有效位改变 50%) 时, $R_M \approx S_M$ 。

以 512×512 大小的灰度静止图像 Lena 为例。将获得的图像数据每四个分为一组, 即 $G = \{x_1, x_2, x_3, x_4\}$, 取秘密信息序列 $\{0, 1, 1, 0\}$ 。然后对该图像分别以嵌入率 10%, 20%, 30%, ..., 90%, 100% 嵌入秘密信息, 分别计算这几种嵌入情况下 R_M 、 S_M 、 R_{-M} 、 S_{-M} 的值, 结果如图 7.5 所示。

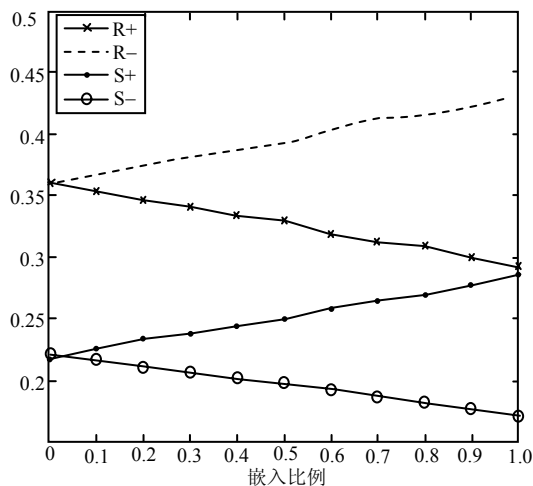


图 7.5 不同嵌入率下 Lena 的 R_M 、 S_M 、 R_{-M} 、 S_{-M} 值

大量的实验表明, R_M 、 S_M 呈线性变化, 而 R_{-M} 、 S_{-M} 则呈现二次多项式曲线(抛物线)。假设一幅图像的嵌入率为 p , 则嵌入率的求解过程如下。

① 对待检测图像的像素进行分组, 分别计算 R_M 、 S_M 、 R_{-M} 、 S_{-M} , 分别记为 $R_M(p/2)$ 、 $S_M(p/2)$ 、 $R_{-M}(p/2)$ 和 $S_{-M}(p/2)$ 。

② 完全随机化待检测图像像素的最低位, 计算得到新的 R_M 、 S_M 、 R_{-M} 、 S_{-M} , 记为 $R_M(1-p/2)$ 、 $S_M(1-p/2)$ 、 $R_{-M}(1-p/2)$ 、 $S_{-M}(1-p/2)$ 。

③ 计算 $d_0 = R_M(p/2) - S_M(p/2)$ 、 $d_1 = R_M(1-p/2) - S_M(1-p/2)$ 、 $d_{-0} = R_{-M}(p/2) - S_{-M}(p/2)$ 、 $d_{-1} = R_{-M}(1-p/2) - S_{-M}(1-p/2)$ 。求解方程 $2(d_0 + d_1)x^2 + 2(d_{-0} - d_{-1} - d_1 - 3d_0)x + d_0 - d_{-0}$ 的根, 取绝对值最小的根 x_m , 那么嵌入比例 $p = x_m / (x_m - 0.5)$ 。

3. DIH 检测方法

DIH (Differential Image Histogram) 算法是由张涛和平西建于 2004 年在软件学报上提出的。他们定义了差分直方图转移系数作为 LSB 平面与图像其余位平面之间弱相关性的度量, 并在此基础上构造区分隐写图像和载体图像的分类器。算法物理意义直观, 实现简单, 计算量小, 效果显著。

记图像 C 在位置 (i, j) 的灰度值为 $C(i, j)$, 则其水平方向上相邻像素差分图像就为 $D(i, j) = C(i, j) - C(i, j+1)$ 。原始图像和载密图像的差分直方图在外形上并没有很大的差异, 但是原始图像的差分直方图在 LSB 平面置反后有明显的变化, 而载密图像的差分直方图在 LSB 平面置反前后却几乎没有任何变化。以大小为 512×512 的灰度图像 Milkdrop 为例, 图 7.6 (b)、(d) 分别是载体图像的直方图以及载体图像 LSB 平面置反后的差分直方图, 而图 7.6 (c)、(e) 是隐写图像 LSB 平面置反前后的差分直方图。DIH 算法就是利用这个统计差异来进行隐写分析的。

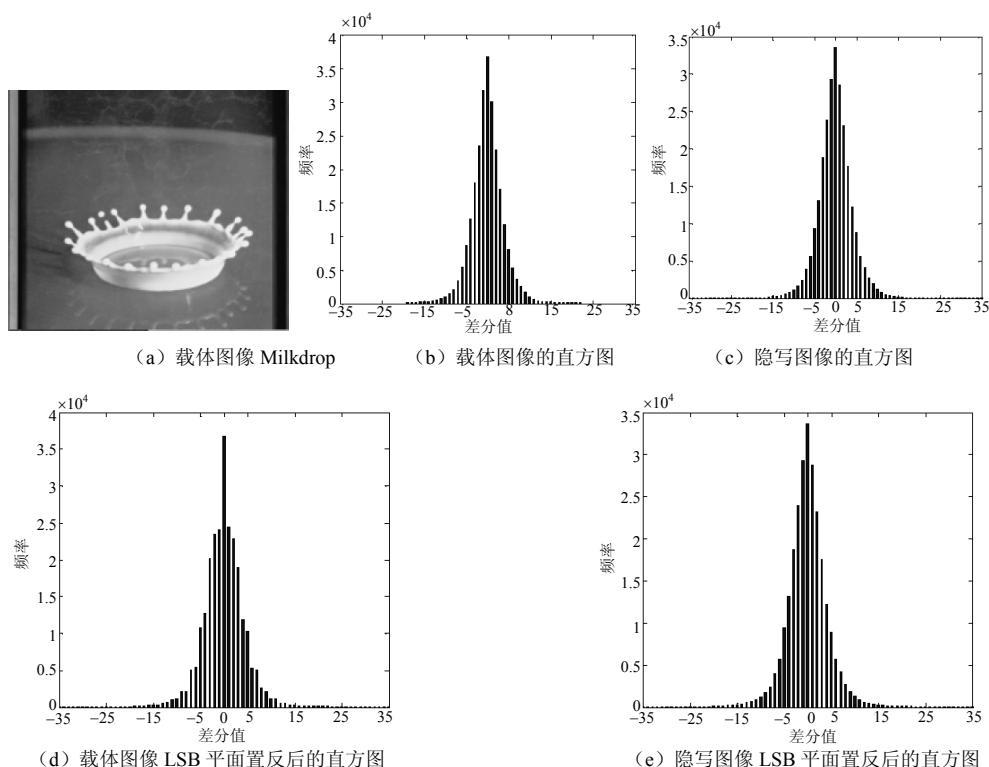


图 7.6 Milkdrop 的差分直方图

记待检图像中所有像素的集合 $S=\{s_1, s_2, s_3, \dots, s_N\}$, $s_{N(j)}$ 表示 s_j 的相邻像素, 定义如下的像素集合

$$H_i = \{s_j \mid s_j - s_{N(j)} = i, j=1, 2, \dots, N\} \quad (7.19)$$

$$G_{2i} = \{s_j \mid \text{int}(s_j / 2) - \text{int}(s_{N(j)} / 2) = i, j=1, 2, \dots, N\} \quad (7.20)$$

其中, $\text{int}(x)$ 代表不大于 x 的最大整数。记待检图像的差分直方图为 $\{h_i \mid h_i = \|H_i\|\}$, 图像 LSB 平面置零后差分直方图为 $\{g_{2i} \mid g_{2i} = \|G_{2i}\|\}$, 其中 $\|A\|$ 代表集合 A 的势。设图像 LSB 平面置反后差分直方图为 f_i , 则 h_i 、 f_i 和 g_i 存在如下关系

$$h_{2i} = f_{2i} = a_{2i,2i} g_{2i} \quad (7.21)$$

$$h_{2i+1} = a_{2i,2i+1} g_{2i} + a_{2i+2,2i+1} g_{2i+2} \quad (7.22)$$

$$f_{2i+1} = a_{2i,2i-1} g_{2i} + a_{2i+2,2i+3} g_{2i+2} \quad (7.23)$$

图 7.7 描述了 h_i 、 f_i 和 g_i 三者之间的转移关系。

由式 (7.21) 可得

$$a_{2i,2i} = h_{2i} / g_{2i} \quad (7.24)$$

若令 $i=0$, 则可得

$$a_{0,0} = h_0 / g_0 \quad (7.25)$$

由于 h_i 和 g_i 近似服从广义拉普拉斯分布, 则

$$a_{0,1} \approx a_{0,-1} \quad (7.26)$$

又因为

$$a_{2i,2i} + a_{2i,2i+1} + a_{2i,2i-1} = 1 \quad (7.27)$$

所以可得

$$a_{0,1} \approx (1 - a_{0,0}) / 2 = (g_0 - h_0) / (2g_0) \quad (7.28)$$

而由式 (7.22) 可得

$$a_{2i,2i-1} = (h_{2i-1} - a_{2i-2,2i-1} g_{2i-2}) / g_{2i} \quad (7.29)$$

由式 (7.27) 得:

$$a_{2i,2i+1} = 1 - a_{2i,2i} - a_{2i,2i-1} \quad (7.30)$$

因此, 根据转移系数计算的初始值式 (7.26)、式 (7.28) 和递推公式 (7.29)、式 (7.30), 就可以求出转移系数。

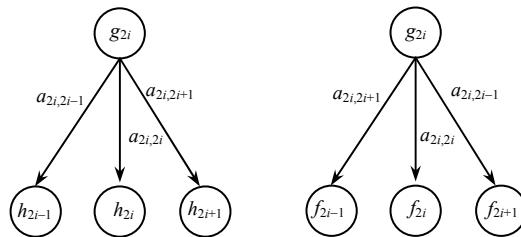


图 7.7 h_i 、 f_i 和 g_i 三者之间的状态转移关系

下面进行秘密信息嵌入率的估计。根据前面求出的转移系数, 记 $a_{2i+2,2i+1} / a_{2i,2i+1}$, $\beta_i = a_{2i+2,2i+3} / a_{2i,2i-1}$, $\gamma_i = g_{2i} / g_{2i+1}$, 假设原始载体满足

$$\alpha_i = \gamma_i \quad (7.31)$$

LSB 全嵌入时满足

$$\alpha_i \approx 1 \quad (7.32)$$

对特定的 i , α_i 的值随嵌入秘密信息长度的增加单调递减, 最后在 LSB 平面全嵌入时减小到 1。DIH 算法选取四个关键点 $P_1=(0, \gamma_i)$, $P_2=(0, \alpha_i)$, $P_3=(1, 1)$, $P_4=(2-p, \beta_i)$ 代入二次多项式 $y=ax^2+bx+c$, 得:

$$\begin{cases} c = \gamma_i \\ ap^2 + bp + c = \alpha_i \\ a(2-p)^2 + b(2-p) + c = \beta_i \\ a + b + c = 1 \end{cases} \quad (7.33)$$

求解方程组 (7.33), 取绝对值最小的根, 即为所求的嵌入率 p 。表 7.1 比较了 RS 和 DIH 算法对 Lena 图像和 Milkdrop 图像的秘密信息嵌入率估计。

表 7.1 RS 和 DIH 算法对两幅经典图像的隐写长度估计

嵌入率 %		0	5	10	20	40	60	80	100
Lena 图像	DIH	-0.773	7.210	17.608	26.736	44.402	63.109	76.145	91.429
	RS	-1.582	4.241	12.383	23.770	44.795	63.899	76.363	100.000
Milkdrop 图像	DIH	-1.182	2.296	6.951	15.244	33.753	55.374	74.671	92.228
	RS	-3.072	0.968	5.087	13.793	31.326	52.822	77.848	100.000

7.3.3 针对 LSB 匹配隐写的专用隐写分析算法

由第 2 章的介绍可知, LSB Matching (也称加减 1) 隐写算法并不会产生值对现象, 因此大多数针对 LSB 替换隐写算法的隐写分析方法不能用于分析 LSB Matching 隐写算法。下面介绍五种典型的针对 LSB 匹配隐写的分析方法。

1. 基于图像直方图特征函数的隐写分析算法

图像直方图描述了不同灰度级像素出现的频率, 能表征图像的一维信息, 秘密信息的嵌入势必对直方图造成影响。一般情况下, 需要嵌入的秘密信息与载体图像是互相独立的, 在此假设下 Harmsem 等^[98]将图像隐写行为模拟为在载体图像中加入随机噪声, 则隐写图像的一维直方图 $h_s(n)$ 可以用载体图像的一维直方图 $h_c(n)$ 和秘密信息分布 $f_m(n)$ 的卷积表示, 即: $h_s(n) = h_c(n) * f_m(n)$ 。傅里叶变换可以化复杂的卷积运算为简单的乘积运算, 因此在傅里叶空间中上式可转换为: $H_s(k) = H_c(k)F_m(k)$ 。对直方图的傅里叶变换即特征函数 (Histogram Characteristic Function, HCF) 定义质心 (Center of Mass, CoM)

$$C[H(k)] = \frac{\sum_{i=0}^{N/2} iH[i]}{\sum_{i=0}^{N/2} H[i]} \quad (7.34)$$

Harmsem 等从理论上证明了 $C(H_s[k]) \leq C(H_c[k])$, 即经过图像隐写后, 图像一维直方图特征函数的质心下降, 可以作为数字图像隐写分析的敏感特征。该特征简记为 Conventional HCF COM。然而, Ker 指出, Conventional HCF COM 特征应用于灰度图像 LSB 匹配隐写分析并不成功。尽管图像隐写后 $C[H]$ 会下降, 但不同载体图像之间的 $C[H]$ 差异很大, 这种差异将覆盖由图像隐写导致的 $C[H]$ 之间的差异 (图 7.8)。Ker 在 HCF COM 的基础上改进性能^[99]: 利用对测试图像下采样可得到一个校正图像 (Calibrated Image), 计算测试图像与校正图像的 HCF COM 可以发现载体图像与隐写图像之间的差

异。通过实验发现,载体图像的四倍下采样图像的 $C(H'_c[k])$ 与载体图像的 $C(H_c[k])$ 几乎相等,而隐写图像的四倍下采样图像的 $C(H'_s[k])$ 与隐写图像的 $C(H_s[k])$ 相差很大,存在关系: $C(H_c[k]) / C(H'_c[k]) < C(H_s[k]) / C(H'_s[k])$ 。此外, Ker 还将 Harmsem 的一维直方图替换为二维直方图。最后,学者们总结得到基于图像直方图特征函数的四种隐写分析特征: Conventional HCF COM、Calibrated HCF COM、Adjacency HCF COM、Calibrated Adjacency HCFCOM。为方便起见,通常将上述四种特征简记为: HCF COMs。这类方法对于 JPEG 解压后的位图文件有一定的效果,但对于未压缩图像会产生较高的虚警率。其原因在于 JPEG 解压后的图像较为平滑,在进行 LSB Matching 嵌入后图像噪声增加明显,而原始的未压缩图像本身纹理就比较复杂。还有学者将 Ker 方法拓展到对差分图像进行校正。

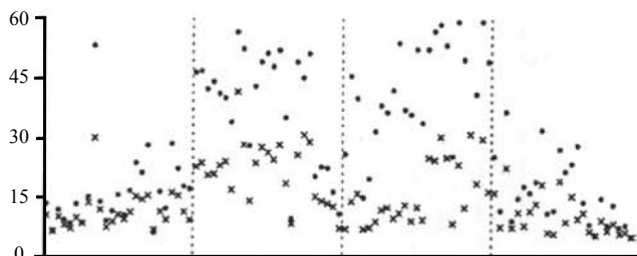


图 7.8 嵌入数据前(圆点)后(叉形)图像的 $C[H]$ 值变化

2. 基于图像直方图局部极值的隐写分析算法

张军等于 2007 年在第九届多媒体信号处理国际会议上指出对于嵌入率为 p 的 LSB 匹配隐写算法,其隐写行为在一维直方图上相当于一个核为 $[p/4, 1-p/2, p/4]$ 的低通滤波,导致图像一维直方图平滑,尤其对局部极值点。定义直方图 $h(n)$ 局部极值与其相邻元素的绝对差的累加和

$$D = \sum_{n^*} |2h(n^*) - h(n^* - 1) - h(n^* + 1)| \quad (7.35)$$

作为隐写分析的特征,其中 n^* 为直方图局部极值点。此外,张军还利用隐写后直方图的局部极大值将变小,而局部极小值将变大的性质,将极大值和极小值分别连接起来,两条边之间的面积用来隐写分析。Cancelli 等于 2008 年在第 15 届图像处理国际会议上扩展了张军提出的图像直方图极值特征。首先,考虑了图像直方图边界的影响。其次,与 Ker 一样,用二维直方图代替了一维直方图。最后隐写分析特征为 10 维,相对原始的直方图局部极值特征,实验性能有了显著改进。该类算法被统称为 ALE。

3. 基于游程长度直方图特征函数的隐写分析算法

游程长度 (Run Length) 又称行程长度,是栅格数据压缩的重要编码方法,能够描述图像灰度关于方向、相邻间隔、变化幅度等信息。游程长度是将图像灰度按照某种方式进行扫描,连续相等的像素点的个数。统计具有相同游程长度 x 的游程出现频数,即得游程长度直方图 $RLH(x)$ 。学者们发现图像经过 LSB 匹配隐写后游程长度直方图会向左移动,即长度较长的游程的数量会减少,而长度较小的游程会增加。这样一来,我们可以用游程长度直方图特征函数质心来刻画 LSB 匹配行为对图像的影响。

4. 基于相关性的隐写分析算法

Liu 等于 2008 年在 Information Sciences 第 178 卷第 1 期提出基于相关性的隐写分析特征,并且指出嵌入率和图像复杂度是影响隐写分析性能的关键因素。LSB 匹配隐写对

图像最低位平面进行修改, 因此图像最低位平面 (LSBP) 与次低位平面 (LSBP2) 之间的相关性会改变, 另外最低位平面自相关性也会发生改变, 基于此提出了基于位平面相关性的多维特征。隐写前后图像直方图发生变化, 可采用图像直方图自相关性进行度量。图像隐写一般模拟为加入噪声, 可以首先对待检测图像进行带阈值的小波降噪, 将去噪后的图像与待检测图像做差, 然后把差值图像的自相关性作为隐写分析的特征。

5. SPAM (Subtractive Pixel Adjacency Model) 隐写分析算法

Pevny 等^[100]指出图像像素的高阶依赖可以使用像素对分布、三像素分布等进行建模, 但是这些建模方法的特征维数随着图像灰度阶以指数方式增加, 并且有些像素组合对于分类还是噪声, 图像内容的多样性也导致很难建立起基于像素组合的模型。Pevny 等在大量图像上统计了像素对的分布情况, 发现大部分相邻像素的像素差值都很小, 并且差值越大出现的频率就越低。上述情况说明自然图像相邻像素的像素值具有很强的连续性, 图像相邻像素相关性很强。实验还发现, 差分图像所携带的信息量与图像数据所携带的信息量几乎相等。因此, Pevny 在八个不同方向的差分图像上进行带阈值的马尔科夫链建模, 这样不仅特征维数低, 差分操作还能消除图像内容多样性对分类性能的影响。若 $I_{i,j}$ 为图像, 在位置 (i,j) 的灰度值, 则水平方向 SPAM 特征提取的步骤如下

第一步: 计算差分图像 D

$$D_{i,j}^{\rightarrow} = I_{i,j} - I_{i,j+1} \quad (7.36)$$

第二步: 在差分图像上建立同方向的带阈值的一阶马尔科夫转移矩阵

$$M_{u,v}^{1st \rightarrow} = \begin{cases} \Pr(D_{i,j+1}^{\rightarrow} = u \mid D_{i,j}^{\rightarrow} = v) & \text{if } u, v \in \{-T, \dots, T\} \\ 0 & \text{if } \Pr(D_{i,j}^{\rightarrow} = v) = 0 \end{cases} \quad (7.37)$$

第三步: 在差分图像上建立同方向的带阈值的二阶马尔科夫转移矩阵

$$M_{u,v}^{2nd \rightarrow} = \begin{cases} \Pr(D_{i,j+2}^{\rightarrow} = u \mid D_{i,j+1}^{\rightarrow} = v, D_{i,j}^{\rightarrow} = w) & \text{if } u, v, w \in \{-T, \dots, T\} \\ 0 & \text{if } \Pr(D_{i,j+1}^{\rightarrow} = v, D_{i,j}^{\rightarrow} = w) = 0 \end{cases} \quad (7.38)$$

第四步: 将同阶的两个水平和两个垂直转移矩阵分别相加得到加和的转移矩阵, 将同阶的四个对角线转移矩阵分别相加得到加和的转移矩阵, 这样做可以降低特征维数

$$\begin{aligned} F_{1,\dots,k}^{1st} &= \frac{1}{4} (M^{1st, \rightarrow} + M^{1st, \leftarrow} + M^{1st, \uparrow} + M^{1st, \downarrow}) \\ F_{k+1,\dots,2k}^{1st} &= \frac{1}{4} (M^{1st, \nearrow} + M^{1st, \swarrow} + M^{1st, \searrow} + M^{1st, \nwarrow}) \\ F_{1,\dots,k}^{2nd} &= \frac{1}{4} (M^{2nd, \rightarrow} + M^{2nd, \leftarrow} + M^{2nd, \uparrow} + M^{2nd, \downarrow}) \\ F_{k+1,\dots,2k}^{2nd} &= \frac{1}{4} (M^{2nd, \nearrow} + M^{2nd, \swarrow} + M^{2nd, \searrow} + M^{2nd, \nwarrow}) \end{aligned} \quad (7.39)$$

SPAM 特征维数对于马尔科夫链的阶和阈值 T 很敏感, 其中一阶 SPAM 特征维数为 $2(2T+1)^2$, 二阶 SPAM 特征维数为 $2(2T+1)^3$ 。为使特征维数适中, 对于一阶特征阈值 T 取 4 和 8, 二阶特征阈值 T 取 3。实验表明阈值为 3 的二阶马尔科夫链特征有最佳检测性能。

7.3.4 针对 JPEG 图像隐写的专用分析算法

针对第 2 章介绍的四种典型 JPEG 隐写算法, 下面介绍相应的典型隐写分析算法。

1. 针对 JSteg 的隐写分析方法

由 Jsteg 算法的嵌入原理可知, 其隐写图像的 DCT 系数直方图中, 仅最低位比特不同的两个 DCT 系数 (除了 0 和 1 之外) 的出现频次趋向一致。根据这一统计特征, 我们可以用卡方检验统计量来判断待检测图像是否嵌入了秘密消息。该方法的关键点在于构造秘密信息理论上的频率分布。对 DCT 系数来说, 最低位为 1 的系数和最低位为 0 的系数构成一个系数对, 成为 POV (Pairs Of Value), 那么嵌入秘密信息前后最低位为 1 和 0 的这对系数之和不变, 也就是这对 POV 的值保持不变。由于嵌入信息服从均匀分布, 那么秘密信息嵌入前后 POV 的个数也保持不变。

以 h_{2i} 表示 DCT 系数值为 $2i$ ($i \neq 0$) 的数目, 令 $n_i = h_{2i}$, 经 Jsteg 算法处理后, n_i 期望值 n'_i 为

$$n'_i = \frac{h_{2i} + h_{2i+1}}{2} \quad (7.40)$$

则统计量

$$r = \sum_{i=1}^k \frac{(n_i - n'_i)^2}{n'_i} \quad (7.41)$$

的分布是渐进自由度为 $(k-1)$ 的 χ^2 分布。 $n_i = n'_i$ 的概率为

$$P = 1 - \frac{1}{2^{\frac{k-1}{2}} \Gamma\left(\frac{k-1}{2}\right)} \int_0^r \exp\left(-\frac{t}{2}\right) t^{\frac{k-1}{2}-1} dt \quad (7.42)$$

其中, k 是 h_{2i} 与 h_{2i+1} 所组成数字对的数量。若概率 P 接近于 1, 说明有秘密消息嵌入, 若 P 值非常小, 甚至接近于 0, 则说明没有秘密消息嵌入。

具体的检测算法流程如下。

- ① 读取整个图像的数据块;
- ② 加载 1% 的图像数据块;
- ③ 根据式 (7.41) 计算 r 值;
- ④ 根据式 (7.42) 计算当前 P 值;
- ⑤ 如果图像未完全加载, 返回至步骤②; 如果图像数据已完全加载, 返回。

上述方法只适用于顺序 JSteg 隐写方法的分析, 而 2002 年 Provos 等提出的通用卡方统计方法^[101]应用范围要广一些, 能对嵌入位置随机排列的 JSteg 隐写算法有效。Lee 等^[102]将卡方检测法进一步深化, 提出了 CA (Category Attack) 检测法和 GCA (Generalized Category Attack) 检测法, 大大改善了卡方检测算法对随机嵌入位置的 JSteg 的检测正确率。Fridrich 等^[103]提出一种移位剪切重压 (Crop and Recompress) 的校正方法, 被广泛应用于多种 JPEG 专用型和通用型隐写分析算法中。这种方法是对 JPEG 图像解压缩后, 剪切前 4 行 4 列, 然后用相同的质量因子再对图像进行 JPEG 压缩, 我们称此图像为校正图像。校正图像的统计特性可以视为是载体图像统计特性的估计。

2. 针对 F5 的隐写分析算法

针对 F5 隐写术, Fridrich 提出了一种典型的针对性隐写分析方法^[104], 它能够检测是否存在隐藏信息, 并能估算隐藏信息长度。该算法分为两步: 确定区分统计量 T , T 与被修改的 DCT 系数总数有关; 确定统计量 T 的基值。

以 $h_{kl}(d)$ 表示在载体图像所有 8×8 的 DCT 矩阵的 (k, l) 位置绝对值为 d 的 DCT 系数总数, $H_{kl}(d)$ 表示在隐写图像相应位置绝对值为 d 的 DCT 系数的数目。若 F5 算法改动了

n 个 DCT 系数, 则一个非“0” DCT 系数被改动的概率为 $\beta=n/P$, 其中 P 为非“0” DCT 系数的总数。因为在 F5 算法中系数的选择是随机的, 所以 $H_{kl}(d)$ 的期望值可表示为

$$H_{kl}(d) = (1-\beta)h_{kl}(d) + \beta h_{kl}(d+1) \quad d > 0 \quad (7.43)$$

$$H_{kl}(0) = h_{kl}(0) + \beta h_{kl}(1) \quad d = 0 \quad (7.44)$$

以 $h'_{kl}(d)$ 表示对载体图像 $h_{kl}(d)$ 的估计, 利用最小均方估计可得 β 的最小值与 $h'_{kl}(d)$ 和 $H_{kl}(d)$ 的关系式

$$\beta_{kl} = \arg \min \{ [H_{kl}(0) - h'_{kl}(0) - \beta h'_{kl}(1)]^2 + [H_{kl}(1) - (1-\beta)h'_{kl}(1) - \beta h'_{kl}(2)]^2 \} \quad (7.45)$$

推导可得

$$\beta_{kl} = \frac{h'_{kl}(1)[H_{kl}(0) - h'_{kl}(0)] + [H_{kl}(1) - h'_{kl}(1)] \cdot [h'_{kl}(2) - h'_{kl}(1)]}{h_{kl}^2(1) + [h'_{kl}(2) - h'_{kl}(1)]^2} \quad (7.46)$$

最终 β 值为所选低频 DCT $(k,l) \in \{(1,2), (2,1), (2,2)\}$ 的平均值

$$\beta = \frac{1}{3}(\beta_{12} + \beta_{22} + \beta_{21}) \quad (7.47)$$

该算法的关键是准确估计载体图像的 $h'_{kl}(d)$, 而要得到准确的 $h'_{kl}(d)$ 必须找到准确的基准图像。获得基准图像分为三步: ① 将隐写图像解压到空域; ② 利用 4 像素在横竖两个方向上对隐写图像进行裁剪; ③ 利用与隐写图像相同的量化矩阵进行压缩。为了得到更为准确的 $h'_{kl}(d)$, 先利用卷积矩阵 \mathbf{B} 对图像进行预处理。 \mathbf{B} 是一个 3×3 矩阵, $B_{22}=1-4e$ 、 $B_{21}=B_{23}=B_{12}=B_{32}=e$ 、其他值为 0。实质上是通过一个低通滤波器, 使获得的低频的 $h'_{kl}(d)$ 更为准确。实验发现, 裁剪图像 (即基准图像) 和载体图像的直方图非常近似, 与隐写图像的直方图在直流上有很大的区别, 从这一点就可以判断哪一个是隐写图像。

采用 $(1, 2^k-1, 1)$ 矩阵编码, 其嵌入效率为 $W(k)$, 令 M 为嵌入消息长度, 编码矩阵参数为 k , 每次改动 DCT 系数可嵌入的消息比特为 $W(k)$

$$W(k) = \frac{2^k}{2^k - 1} \cdot k \quad (7.48)$$

定义一个 DCT 系数可能产生收缩的概率为 $P_s=h(1)/P$, 得出对 M 的估计

$$M = \frac{2^k}{2^k - 1} \cdot k \cdot n \cdot (1 - P_s) = \frac{2^k}{2^k - 1} \cdot k \cdot \beta \cdot [P - h(1)] \quad (7.49)$$

其中

$$P = \sum_{i \geq 0} h(i) \approx \sum_{i \geq 0} \sum_{\substack{k,l=1 \\ k+l > 2}}^8 h'_{kl}(i) \quad (7.50)$$

综上, 可以得到 F5 隐写分析算法流程如下: 算法输入为隐写图像, 输出为秘密信息长度 M 。则 F5 隐写分析算法如下。

- (1) 由式 (7.50) 求出所有非零 AC 系数的总数;
- (2) 由式 (7.47) 估计 β 的值;
- (3) 计算 $n=\beta \times P$, 由 $n=2^k-1$ 求出 k ;
- (4) $h(1)$ 是 AC DCT 系数绝对值为 1 的数量, 由量化 DCT 系数可计算 $h(1)$;
- (5) 将以上计算结果代入式 (7.49), 可得出秘密信息长度 M 的估计值。

3. 针对 OutGuess 的隐写分析算法

OutGuess 调整了隐写图像的直方图, 使其和载体图像完全一致, 所以一般的直方图攻击方法不能用来检测 OutGuess 隐写图像。Fridrich 等人通过大量实验发现将 JPEG 图

像解压缩到空间域然后做图像边缘少量切割后, 其边界不连续点的分布特性与原始载体图像相接近, 由此提出了针对 OutGuess 的隐写分析算法^[105], 该算法还可估计嵌入的秘密信息长度。其基本步骤如下。

(1) 首先将待检测图像解压缩到空间域, 然后计算空间域的边界不连续点, 记为 $B_s(0)$;

(2) 在待检测图像中用 OutGuess 隐写方法嵌入最大长度的秘密信息, 解压缩, 计算其图像空间域边界不连续点 $B_s(1)$, 记 $S=B_s(1)-B_s(0)$;

(3) 对步骤 1 中的空间域图像, 在四个方向均裁剪掉 4 个像素, 然后用与原图像相同量化表压缩。再将压缩后的 JPEG 图像解压到空间域, 并计算边界不连续点 $B(0)$;

(4) 对上一步得到的 JPEG 图像中用 OutGuess 隐写算法嵌入最大长度的秘密信息然后解压缩, 计算其空间域边界不连续点 $B_1(0)$;

(5) 将最大长度的秘密信息再嵌入到上一步得到的载密图像中, 计算图像空间域边界不连续点 $B_1(1)$;

(6) 计算秘密信息的嵌入率如下

$$p = \frac{S_0 - S}{S_0 - S_1} \quad (7.51)$$

其中 $S_0=B(1)-B(0)$, 将其作为原始载体图像的特性; $S=B_s(1)-B_s(0)$ 代表待检测图像的特性, $S_1=B_1(1)-B_1(0)$ 是包含最大长度秘密信息的载密图像的特性; S 应该在 S_1 和 S_0 之间。设 p 为待检测图像中秘密信息嵌入率 (p 的值在 0 和 1 之间), 则有 $S=S_0-p(S_0-S_1)$ 。

4. 针对 MB 的隐写分析算法

第 2 章中给出的针对 JPEG 图像的 MB 隐写算法的基本原理可概括如下: 为了使隐写图像和载体图像的 AC 系数具有相同的广义柯西分布的模型 $f(u, p, s)$, 把秘密消息根据该模型进行算术解码后再嵌入到载体图像的非零 AC 系数的 LSB 上。秘密消息嵌入前后, AC 系数的低精度直方图 $b^{(ij)}$ 保持不变, 那么据此可估计出模型 $f(u, p, s)$, 因此只要将隐写图像的非零 AC 系数的 LSB 根据该模型进行算术编码后可还原出秘密消息。

尽管自然图像的 AC 系数直方图可以较好地近似为柯西分布, 但仍有一部分 AC 系数值的出现频次不符合该分布。由于采用 MB 算法嵌入消息后将使出现频次不符合柯西分布的 AC 系数值的数目减少, 从而造成隐写图像和载体图像之间的特征差异。从这一角度出发, 可以提出针对 MB 的检测算法。

对于载体图像, 可以根据它的 $b^{(ij)}$ 估计出 (i, j) 位置上 AC 系数的期望模型 $f(u, p, s)$ 。采用 MB 嵌入秘密消息后, $b^{(ij)}$ 不变, 所以隐写图像的 AC 系数和载体图像的 AC 系数保持了相同的期望模型, 即

$$p(u) = \frac{p-1}{2s} \left(\left| \frac{u}{s} \right| + 1 \right)^{-p} \quad (7.52)$$

其分布函数为

$$D(u) = \begin{cases} 0.5(|u/s| + 1)^{1-p} & u \leq 0 \\ 1 - 0.5(|u/s| + 1)^{1-p} & u > 0 \end{cases} \quad (7.53)$$

量化后的 DCT 系数为整数, 可以用 U 表示, 其出现概率可以由下式计算

$$P_d(U = k) = D(k + 0.5) - D(k - 0.5) \quad (7.54)$$

根据极大似然法由 $b^{(ij)}$ 估计出 $f(u, p, s)$ 模型的参数 p 和 s 。

隐写图像的高精度直方图 $h^{(i,j)}$ 服从二项分布

$$\begin{cases} h_{2k-1}^{(i,j)} \sim B(b_k^{(i,j)}, p_k^{(i,j)}) \\ h_{2k}^{(i,j)} \sim B(b_k^{(i,j)}, 1 - p_k^{(i,j)}) \end{cases} \quad (7.55)$$

其期望值

$$\begin{cases} \tilde{h}_{2k-1}^{(i,j)} = E[B(b_k^{(i,j)}, p_k^{(i,j)})] = b_k^{(i,j)} p_k^{(i,j)} \\ \tilde{h}_{2k}^{(i,j)} = E[B(b_k^{(i,j)}, 1 - p_k^{(i,j)})] = b_k^{(i,j)} (1 - p_k^{(i,j)}) \end{cases} \quad (7.56)$$

其中, $p_k^{(i,j)}$ 可由下式计算出

$$p_k^{(i,j)} = \frac{f(2k-1, p, s)}{f(2k-1, p, s) + f(2k, p, s)} \quad (7.57)$$

秘密消息根据 $p_k^{(i,j)}$ 进行算术解码后嵌入到载体图像后, 会使得载密图像的高精度直方图 $h^{(i,j)}$ 更加接近式 (7.56) 所示期望值。

为了衡量隐写图像的直方图和期望值的接近程度, 可采用 χ^2 检验法。检测一幅图像是否采用 MB 算法嵌入了秘密消息, 如果对该图像的所有 DCT 块中 63 个位置的 AC 系数直方图的所有的 $b^{(ij)}$ 作 χ^2 检验, 则计算量会过于庞大, 因此, 只取 $b_1^{(ij)}$ 和 $b_{-1}^{(ij)}$ 作 χ^2 检验, 也就是只需做 126 次 χ^2 检验。把 126 次检验中不符合期望分布的数目 n 作为该幅图像的特征值。根据 MB 算法原理可知, 与载体图像相比, 隐写图像的特征值 n 较小, 则可以利用该特征值对隐写图像进行检测。

7.4 图像通用隐写分析

7.4.1 图像通用隐写分析基本思想

通用隐写分析, 为一类具有自学习能力的隐写分析方法, 其攻击对象包括了空域与变换域隐写术, 是隐写分析领域不可忽视的部分。其基本思路是: 通过大量载体图像和隐秘图像的训练, 构造分类器, 可以检测出由不同隐写算法嵌入的秘密消息。这类方法一般围绕嵌入秘密消息前后图像的总体、局部、相关等特征的变化找出一些具有区分能力的统计量构成特征矢量, 经过特征选择以后, 利用回归分析、Fisher 线性分类准则、神经网络或者支持向量机等方法从训练数据中构造检测模型, 与此同时也要找出与之相适应的判决阈值, 该检测模型被用作隐写图像和载体图像的分类器。其检测对象一般是针对多类隐藏方法或多种隐藏工具, 不管隐写算法作用在空域还是变换域, 它不关注或剖析具体隐写算法的细节, 这也是“通用”的含义所在。但是, 通用特征的选取和阈值的确定非常困难, 而且复杂度偏高, 实用性不强, 准确性较低, 无法控制虚警率和漏报率。尽管通用性隐写分析方法的检测率普遍没有针对性隐写分析高, 但是通用隐写分析方法具有较好的适应性。对于新的隐写方法, 针对性的分析方法需要根据其隐藏机制重新设计特征, 而通用隐写分析方法仅需要重新训练分类器即可。

通常说来, 通用隐写分析过程可以分为以下 5 个步骤。

(1) 图像预处理。如灰度转化、裁剪大小、小波变换、信息嵌入、确定分布模型等。

(2) 特征提取。通用隐写分析方法的好坏主要取决于所选取特征的好坏。良好的特征应当包含数据嵌入导致图像变化的信息, 而不含图像本身内容, 且图像的特征应该随嵌入数据量大小的变化而变化, 若嵌入的数据量越大, 特征的分类也越容易。提取的特征有很多种, 如空域特征、DCT 域特征、小波域高阶统计特征等。

(3) 分类器选择。根据提取的特征选择合适的分类器，很多的隐写分析算法选择**支持向量机** (Support Vector Machine, SVM)，因为它在非线性分类上具有显著优势。**Fisher 线性分类器** (Fisher Linear Discriminant, FLD) 使用得也相当广泛，因为它简单并有效。另外的一些技术，如多元回归分析、贝叶斯分类器和神经网络等，也可以用来分类。

(4) 分类器训练。利用已知类别的图像对分类器进行训练，得出相关参数，并进行参数调节。

(5) 分类。信息隐写的通用检测问题可看作分为两类的分类问题。设定门限值，利用训练好的分类器对待测试图像进行分类，可对图像中是否含有秘密信息做出判决。

通用隐写分析方法的基本原理的框架图如图 7.9 所示。需要指出，有的方法在提取特征后，先将特征值进行聚类操作，然后进行分类，测试图像可能是实验者搜集的样本，也可能是来自其他途径的待检测图像。

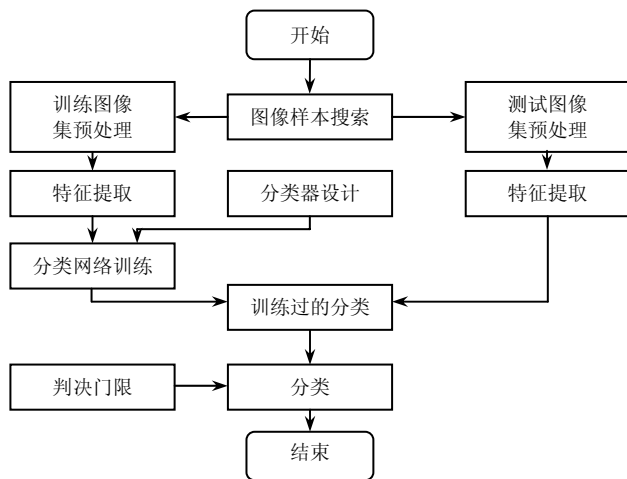


图 7.9 图像通用隐写分析基本原理框架

7.4.2 支持向量机分类技术

1. 支持向量机基本思想

支持向量机是 Vapnik 等根据**统计学习理论** (Statistical Learning Theory, SLT) 提出的一种新的机器学习方法，在解决小样本、非线性及高维模式识别问题中表现出许多特有的优势，已经在模式识别、函数逼近和概率密度估计等方面取得了良好的效果。图像隐写分析领域是支持向量机分类应用的重要领域。本书后面的例子就是利用了支持向量机的分类功能，即将输入图像分为两类：一类是没有经隐写处理的载体图像类，另一类是经过隐写技术处理的隐写图像类。

支持向量机从本质上讲是一种前向神经网络，根据结构风险最小化准则，在使训练样本分类误差极小化的前提下，尽量提高分类器的泛化推广能力。从实施的角度，训练支持向量机的核心思想等价于求解一个线性约束的二次规划问题，从而构造一个超平面作为决策平面，使得特征空间中两类模式之间的距离最大，而且它能保证得到的解为全局最优解。以图 7.10 所示的二维线性可分情况为例，SVM 的基本原理是利用分类超平面尽可能多的将空间中两类样本点正确的分离，同时使得分开的两类样本点距离分类器最远，图中空圆点和空方点分别代表了两类样本， H 是分类线， H_1 、 H_2 分别是过各类中

离分类线最近的样本且平行于分类线的直线，而分类间隔就是指它们之间的距离，支持向量则指的是 H_1 、 H_2 上的训练样本点（实圆点和实方点）。如果分类线既能将两类正确分开，又能使分类间隔最大，则被认为是最优分类线。若在高维空间，则得到的是最优分类面。下面分为线性和非线性两种情况进行讨论。

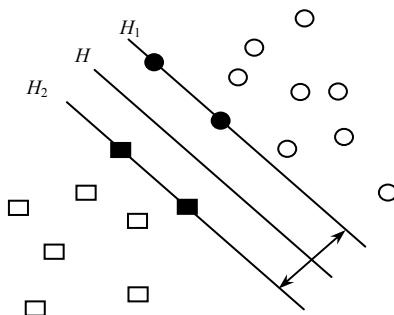


图 7.10 支持向量机最优分类超平面示意图

2. 线性支持向量机分类器

SVM 方法是从线性可分情况下的**最优分类面**（Optimal Hyperplane）提出的。所谓最优分类面就是要求分类面不但能将两类样本无错误的分开，而且要使两类之间的距离最大。设样本集为 $\{\mathbf{x}_i, y_i\}$, $i=1, 2, \dots, n$, $\mathbf{x}_i \in \mathbb{R}^d$ 为 d 维数据矢量, $y_i \in \{+1, -1\}$ 是类别标号。 d 维空间中线性判别函数的一般形式为 $g(\mathbf{x})=\mathbf{w} \cdot \mathbf{x}+\mathbf{b}$, 其分类面方程为

$$\mathbf{w} \cdot \mathbf{x}+\mathbf{b}=0 \quad (7.58)$$

在线行可分情况下，可将判别函数进行归一化，使两类所有样本都满足 $|g(\mathbf{x})| \geq 1$ ，也就是说使离分类面最近的样本的 $|g(\mathbf{x})|=1$ ，这样分类间隔就等于 $2/\|\mathbf{w}\|$ ，故间隔最大等价于使 $\|\mathbf{w}\|$ （或 $\|\mathbf{w}\|^2$ ）最小。从而，要求分类面对所有样本 $\{\mathbf{x}_i, y_i\}$, $i=1, 2, \dots, n$ 正确分类，就是要求其满足

$$y_i[(\mathbf{w} \cdot \mathbf{x}_i)+\mathbf{b}]-1 \geq 0, (i=1, 2, \dots, n) \quad (7.59)$$

因此，满足上述条件且使 $\|\mathbf{w}\|^2$ 最小的分类面就是最优分类面。在这两类样本中，离分类面最近的点且平行于最优分类面的超平面上的训练样本就是使式（7.59）中等号成立的那些样本，它们叫做**支持向量**（Support Vectors）。根据上面的讨论，最优分类面问题可以表示成如下的约束优化问题，即在式（7.59）的约束下，求如下函数的最小值

$$\varphi(\mathbf{w})=0.5\|\mathbf{w}\|^2=0.5\mathbf{w} \cdot \mathbf{w} \quad (7.60)$$

这是一个二次规划问题，可定义以下的拉格朗日函数

$$L(\mathbf{w}, \mathbf{b}, \mathbf{a})=0.5\|\mathbf{w}\|^2-\sum_{i=1}^n a_i \cdot\left\{y_i \cdot[(\mathbf{w} \cdot \mathbf{x}_i)+\mathbf{b}]-1\right\} \quad (7.61)$$

其中， $a_i \geq 0$ 为拉格朗日乘子。求式（7.60）的极小值就是对 \mathbf{w} 和 \mathbf{b} 求拉格朗日函数的极小值。求 L 对 \mathbf{w} 和 \mathbf{b} 的偏微分，并令其等于 0，可得

$$\mathbf{w}=\sum_{i=1}^n a_i y_i \mathbf{x}_i, \quad \sum_{i=1}^n a_i y_i=0, \quad i=1, 2, \dots, n \quad (7.62)$$

并将问题转化为在约束条件 $\sum_{i=1}^n a_i y_i=0$, $a_i \geq 0$, $i=1, 2, \dots, n$ 之下，对 \mathbf{a} 求下式的最大值

$$W(\mathbf{a}) = \sum_{i=1}^n a_i - 0.5 \sum_{i=1}^n \sum_{j=1}^n a_i a_j y_i y_j \cdot (\mathbf{x}_i \cdot \mathbf{x}_j) \quad (7.63)$$

对于线性不可分的样本集，人们希望误分类的点数最小，为此可在式 (7.59) 中引入松弛变量 $\xi_i \geq 0$ ，即

$$y_i \cdot [(\mathbf{w} \cdot \mathbf{x}_i) + \mathbf{b}] - 1 + \xi_i \geq 0, \quad i=1, 2, \dots, n \quad (7.64)$$

对于给定的惩罚参数 C ，求出使下式取极小值的 \mathbf{w} 和 \mathbf{b}

$$\varphi(\mathbf{w}, \boldsymbol{\xi}) = 0.5 \mathbf{w} \cdot \mathbf{w} + C \cdot \sum_{i=1}^n \xi_i \quad (7.65)$$

其中， C 值的大小与对错误分类的惩罚程度成正比。对于这一优化问题，可定义如下拉格朗日函数

$$L(\mathbf{w}, \mathbf{b}, \mathbf{a}, \boldsymbol{\beta}, \boldsymbol{\xi}) = 0.5 \|\mathbf{w}\|^2 + C \cdot \sum_{i=1}^n \xi_i - \sum_{i=1}^n a_i \cdot \{y_i \cdot [(\mathbf{w} \cdot \mathbf{x}_i) + \mathbf{b}] - 1 + \xi_i\} - \sum_{i=1}^n \beta_i \cdot \xi_i \quad (7.66)$$

其中， $a_i \geq 0$ ， $\beta_i \geq 0$ 为拉格朗日乘子。求 L 对 \mathbf{w} 和 \mathbf{b} 的偏微分，并令其等于 0，可得

$$\mathbf{w} = \sum_{i=1}^n a_i y_i \mathbf{x}_i, \quad \sum_{i=1}^n a_i y_i = 0, \quad C - a_i - \beta_i = 0, \quad i=1, 2, \dots, n \quad (7.67)$$

并将上述问题转化为在下式的约束条件下，对 \mathbf{a} 求式 (7.63) 的最大值：

$$\sum_{i=1}^n a_i y_i = 0, \quad 0 \leq a_i \leq C, \quad i=1, 2, \dots, n \quad (7.68)$$

式 (7.66) 的最优化求解得到的 a_i 可能是：① $a_i = 0$ ；② $0 < a_i < C$ ；③ $a_i = C$ ，后面两种情况对应的 \mathbf{x}_i 就是支持向量。这样一来，由式 (7.67) 可知，只有支持向量才对 \mathbf{w} 有贡献，也就是对最优超平面和判别函数都有利，于是 \mathbf{w} 可表示为

$$\mathbf{w} = \sum_{\text{Support Vectors}} a_i y_i \mathbf{x}_i \quad (7.69)$$

即最优分类面的权系数向量是支持向量的线性组合，此处所对应的学习方法被称作支持向量机。而在支持向量中，第③种情况对应的 \mathbf{x}_i 叫做边界支持向量，也就是错分的训练样本点；第②种情况对应的 \mathbf{x}_i 称作标准支持向量。而由 Karush-Kuhn-Tucker 条件可知，对应最优点的情况，拉格朗日乘子与约束的积是 0，如下式所示

$$a_i \cdot \{y_i \cdot [(\mathbf{w} \cdot \mathbf{x}_i) + \mathbf{b}] - 1 + \xi_i\} = 0, \quad \beta_i \cdot \xi_i = 0, \quad \forall i \quad (7.70)$$

对于 $0 < a_i < C$ 的标准支持向量 \mathbf{x}_i 来说，可由 $C - a_i - \beta_i = 0$ 得出 $\beta_i > 0$ ，通过式 (7.70) 得出 $\xi_i = 0$ ，故对任何一个标准支持向量 \mathbf{x}_i ，符合

$$y_i \cdot [(\mathbf{w} \cdot \mathbf{x}_i) + \mathbf{b}] = 1 \quad (7.71)$$

即 $y_i = \mathbf{w} \cdot \mathbf{x}_i + \mathbf{b}$ ，从而可得

$$\mathbf{b} = y_i - \sum_{j=1}^n a_j y_j (\mathbf{x}_j \cdot \mathbf{x}_i) \quad (7.72)$$

其中 \mathbf{x}_i 为标准支持向量。对所有标准支持向量分别计算 \mathbf{b} 值，求其平均值即可。

在得到最优解 \mathbf{w}^* 、 a_i^* 和 \mathbf{b}^* 后，对于给定的未知样本 \mathbf{x} ，只需计算下式的最优分类函数即可判定 \mathbf{x} 所属的分类

$$f(\mathbf{x}) = \text{sgn}\{(\mathbf{w}^* \cdot \mathbf{x}) + \mathbf{b}^*\} = \text{sgn}\left\{\sum_{i=1}^n a_i^* \cdot y_i \cdot (\mathbf{x}_i \cdot \mathbf{x}) + \mathbf{b}^*\right\} \quad (7.73)$$

其中, sgn 为符号函数。

3. 非线性支持向量机分类器

当训练集为非线性时, 由非线性函数 $\phi(\mathbf{x})$ 将训练集数据 \mathbf{x} 映射到一个高维线性特征空间, 在该高维线性空间中, 构造最优分类超平面, 从而得到分类器判别函数。因此, 在非线性的情况下, 分类超平面为

$$\mathbf{w} \cdot \phi(\mathbf{x}) + b = 0 \quad (7.74)$$

判别函数为

$$f(\mathbf{x}) = \text{sgn}\{[\mathbf{w} \cdot \phi(\mathbf{x})] + b\} \quad (7.75)$$

最优分类超平面问题描述为:

$$\begin{cases} \min \{0.5 \mathbf{w} \cdot \mathbf{w} + C \cdot \sum_{i=1}^n \xi_i\} \\ y_i [\mathbf{w} \cdot \phi(\mathbf{x}_i) + b] - 1 + \xi_i \geq 0 \end{cases} \quad (7.76)$$

其中 $\xi_i \geq 0, i=1, 2, 3, \dots, n$ 。类似上面的线性情况, 可得出判别函数为

$$f(\mathbf{x}) = \text{sgn}\left\{\sum_{i=1}^n a_i^* \cdot y_i \cdot K(\mathbf{x}_i, \mathbf{x}) + b^*\right\} \quad (7.77)$$

其中, $K(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x}) \cdot \phi(\mathbf{x}')$ 称作核函数。

尽管可以通过非线性函数将样本数据映射到较高维空间, 并可在特征空间中构造最优分类超平面, 但在求解最优化问题和计算判别函数时并不必显式计算该非线性函数, 而只需要计算核函数, 因此避免了特征空间维数灾难的问题。同时支持向量机的推广性与变换空间的维数无关, 只要能够适当地选择一种内积定义, 构造一个支持向量数相对较少的最优或者广义最优分类, 就可以得出较好的推广性。

由泛函的相关理论可知, 如果一个函数满足 Mercer 条件那么就认为该函数是关于某个变换的核函数, 简化的 Mercer 条件如下式所示

$$\int_{\mathcal{X} \times \mathcal{X}} K(\mathbf{x}, \mathbf{x}^*) f(\mathbf{x}) f(\mathbf{x}^*) d\mathbf{x} d\mathbf{x}^* \geq 0, \quad f \in L_2(\mathcal{X}) \quad (7.78)$$

其中 $K(\mathbf{x}, \mathbf{x}^*)$ 是平方可积空间中的函数, $f(\mathbf{x})$ 是平方可积空间中的任意函数, $L_2(\mathcal{X})$ 表示平方可积空间。若公式 (7.78) 成立, 那么 $K(\mathbf{x}, \mathbf{x}^*)$ 就是核函数。

常见的核函数主要有以下几种。

(1) 线性核函数

$$K(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot \mathbf{y} \quad (7.79)$$

(2) 多项式核函数

$$K(\mathbf{x}, \mathbf{y}) = [(\mathbf{x} \cdot \mathbf{y}) + \theta]^l \quad (7.80)$$

(3) 高斯核函数 (径向基核函数 RBF)

$$K(\mathbf{x}, \mathbf{y}) = \exp[-\|\mathbf{x} - \mathbf{y}\|^2 / (2\sigma^2)] \quad (7.81)$$

(4) 二层神经网络内积函数

$$K(\mathbf{x}, \mathbf{y}) = \tanh(u \cdot (\mathbf{x} \cdot \mathbf{y}) + c) \quad (7.82)$$

核函数又分别对应了不同的学习机器。一般情况下, 只需要改变核函数的形式就可以构造出不同的学习机器, 而其他形式的核函数还有很多。由于高斯核函数简单实用且只有一个参数, 人们通常选用高斯核函数。

综上, 支持向量机的基本思想可以概括为: 首先通过非线性变换将输入空间变换到

一个高维空间,然后在这个新空间中求取最优线性分类面,而这种非线性变换是通过定义适当的核函数实现的。

7.4.3 图像通用隐写分析算法概述

通用隐写分析方法最早由 Avcibas 和 Farid 提出,国内外很多专家在此领域进行了深入的研究。下面,分类归纳一些主要的通用隐写分析算法。

1. 基于图像质量度量的通用隐写分析算法

Avcibas 等^[106]提出了最早的一种通用方法,称为基于**图像质量度量**(IQMs, Image Quality Metrics)的隐写分析方法。这些度量分为基于图像相关性、图像边缘、像素差分、频谱和 HVS(人类视觉系统)等。对这 26 种不同的度量进行分析,将这些度量用来预测压缩、模糊和加噪,结果发现基于 HVS、相位频谱以及多分辨均方误差这三类度量具有很好的区分能力。Avcibas 等采用了如下质量度量作为特征。

- (1) 基于图像均方误差的质量度量;
- (2) 基于图像多分辨率距离的质量度量;
- (3) 基于结构内容的质量度量;
- (4) 基于图像互相关性的质量度量;
- (5) 基于图像加权频谱距离的质量度量;
- (6) 基于图像中值块加权频谱距离的质量度量;
- (7) 基于标准化绝对 HVS 误差的质量度量;
- (8) 基于 HVS 均方误差的质量度量;
- (9) 基于梯度度量的质量度量。

基于 IQM 的检测算法首先对训练图像集中的原始图像和隐写图像进行滤波,从滤波后的图像中提取 n 维特征,再将这些 IQM 值回归为输出值,如果输出值大于门限 0,则认为图像被嵌入了信息,否则认为没有被嵌入信息。该方法是最早的通用隐写分析方法,但检测效果一般。

2. 基于小波高阶统计量的通用隐写分析方法

基于小波高阶统计量的通用隐写分析方法由 Farid 最先提出^[107],该方法利用**正交镜像滤波器**(Quadrature Mirror Filter, QMF)对图像进行分解,将图像分成垂直、水平、对角和低通 4 个子带,然后对低通子带进行递归分解操作,提取各个小波子带系数及其线性预测误差的前四阶统计矩(均值、方差、偏度和峰度)来进行统计分析,总共得到了 72 维特征量。作者先后用 Fisher 线性分类器和支持向量机 SVM 分类器进行了实验,并指出利用非线性 SVM 得到的检测率较高。而小波高阶统计量的方法也取得了一定的检测效果。Farid 等提出的对图像进行小波分解,以及采用高阶矩作为图像特征方法,被以后的很多通用隐写分析算法所采用。黄聪等学者将小波直方图频域矩的通用隐写方法推广到 JPEG 图像中,使之成为一个不仅能处理 JPEG 图像并且可以处理 BMP 图像的高效系统。针对 BMP 图像上较流行的 QIM、SS(Spread Spectrum)和 LSB 等隐写方法,在 Coreldraw 图像库上进行了大量的隐写分析实验,检测效果良好;而且该检测方法对于当前公认安全性较高的 JPEG 图像隐写方法 MB、Outguess 和 F5 等也有很好的检测性能。

3. 基于直方图特征函数质心的通用隐写分析方法

Harmsen 和 Pearlman^[98]将信息隐写的过程建模为加性噪声过程,在假设嵌入信息独

立于原始图像的条件下, 隐写图像的直方图可看作噪声概率质量函数 (Probability Mass Function, PMF) 和原始图像的直方图的卷积, 在频域上, 这种卷积被视为直方图特征函数 (Histogram Characteristic Function, HCF) 和噪声特征函数的乘积。LSB 隐写、扩频隐写及 DCT 域隐写方法均可纳入该模型, 隐藏信息后, HCF 的质心将降低, 这种降低被用于隐写检测。实验中的分类器采用了贝叶斯分类器。该方法对于低噪声的彩色图像具有很好的分析效果, 但是对于从扫描仪或数字照相机中获得的原始的且未经压缩的灰度图像, 检测效果明显偏低。该方法还首次提出了信息隐藏的加性噪声模型并且将特征函数应用到隐写分析领域, 方便后继的学者进行研究。邓艺等人于 2010 年在中国科学院研究生院学报第 27 卷第 3 期提出了一种基于质量因子和特征函数的 JPEG 图像通用隐写分析方法, 该方法通过分析 JPEG 图像小波域和 DCT 域的量化隐写模型和噪声模型, 得到嵌入噪声和量化噪声对图像的作用原理, 且通过直方图特征函数加以区别这两种噪声对图像造成的影响, 并为每个图像分类单独训练支持向量机分类器。

4. 基于 DCT 特征的 JPEG 图像通用隐写分析方法

Fridrich 等^[108]提出一种基于 DCT 特征的 JPEG 图像通用隐写分析方法, 该方法直接从 JPEG 图像的 DCT 域中提取 23 个特征, 这 23 维特征量分为 6 类: ① 3 个 DCT 系数的共生矩阵; ② 2 个 DCT 块的不连续性特征; ③ 1 个 DCT 系数的方差; ④ 11 个双重 DCT 系数直方图; ⑤ 5 个特定位置的 DCT 系数直方图; ⑥ 1 个全局 DCT 系数直方图。在提取的 DCT 特征中, 前 3 类为二阶特征, 后 3 类为一阶特征, 它们较全面地反映了信息嵌入后 JPEG 图像中 DCT 系数可能变化的统计量。Fridrich 把 JPEG 图像简单地分解到空域, 然后沿空域的边缘去掉 4 个像素, 再利用原始量化表把图像重压缩为 JPEG 图像, 分别提取原图像和重压缩图像的特征, 接着两者相减提取范数, 他接着利用特征校准技术对得到的特征进行处理, 使得特征更加敏感。Fridrich 的 DCT 特征方法对当前流行的 JPEG 类数据嵌入算法, 如 F5、Outguess、Jsteg、Steghide、JPhide 和 MB 算法均取得了很高的检测效果。但是 DCT 特征的提取方法十分复杂, 运算量大, 并且只能对 JPEG 图像进行检测。徐志杰等人于 2010 年在信息与电子工程第 8 卷第 1 期提出了一种基于离散余弦变换系数统计特性的 JPEG 图像隐写分析方法。该方法对 JPEG 图像 DCT 系数的统计特性进行研究, 分别提取了 DCT 分块特性参数、共生矩阵参数、差分直方图参数共 8 维特征向量, 并用最小二乘支持向量机 (LS-SVM) 分类器对待测图像进行分类, 最后检测出隐写图像, 该方法能够有效检测各类 JPEG 图像隐写算法。

5. 基于图像噪声统计矩的通用隐写分析

Holotyak 等^[109]通过对图像进行正交小波分解, 估计出特定小波分解层的噪声分量和该层小波分解的能量并提取噪声的高阶概率密度函数矩作为特征。Goljan 等^[110]在 Holotyak 等方法的基础上做出了一些改进, 他们考虑了图像的非平衡特性, 并使用维纳滤波器对图像去噪, 最后提取噪声的绝对值概率密度函数中心矩作为特征, 结果表明该方法对未压缩的原始图像和灰度图像的隐写检测效果要比 JPEG 图像和 RGB 图像的检测效果差。另外, Holotyak 等^[109]给出的含密噪声的概念被后来学者所接受。

6. 基于小波系数直方图特征函数统计矩的通用隐写分析方法

Xuan 等^[111]在 Farid^[107]及 Harmsen 等^[98]算法的基础上, 对图像进行两级 Haar 小波分解, 提取小波系数的 3 阶绝对值特征函数矩作为特征。随后, Shi 等提出了改进版本^[112], 在 Xuan 等特征的基础上, 对图像进行两级 Haar 小波分解, 得到 9 个子带, 对每一

个子带系数的直方图计算其特征函数,提取一阶矩和二阶矩为特征,把得到的 18 维特征采用贝叶斯分类器进行分类。结果表明:他们的方法对 Cox 提出的非盲扩频水印算法的正确检测率为 79%,对 LSB 替换隐写的正确检测率高达 91%,而且基于特征函数矩的特征要优于基于概率密度函数矩的特征。随后 Wang 等^[113]从图像子带分解、特征选择、特征评估三方面做了研究,指出特征函数矩优于概率密度函数矩,并在理论上给出证明。并且 Wang 等将模式识别和机器学习中的特征降维技术应用到图像隐写分析中,对提取的特征进行进一步处理,以提高检测算法的性能,结果表明:Wang 等的特征要优于 Farid 等和 Xuan 等的特征,不仅对各种 LSB、F5、Outguess 等隐写算法有效,而且对几种扩频隐写算法也得到了高检测率。不过,对于含纹理较多的图像,检测效果要稍差。

7. 基于二元相似性度量的通用隐写分析方法

Avcibas 等^[114]最早提出的第二种方法是二元相似性度量 BSM (Binary Similarity Measures),其中特征矢量取自图像空域。他们认为在信息嵌入后,相邻位平面之间的相关性减少。Avcibas 等人对图像的第 7 层位平面和第 8 层位平面进行了研究,得到了三类特征矢量:熵矢量、计算相似性差别方图和一系列基于相邻权重掩饰的矢量,并使用线性回归分类器进行分类。但是,实验结果表明该方法的检测效果不甚理想。

8. 基于共生矩阵的通用隐写分析方法

Chen 等^[115]将共生矩阵用于图像的通用检测中,其基本思想是:从经验矩阵的投影直方图 (Projection Histogram, PH) 中抽取两种特征:PH 的矩以及 PH 的特征函数矩,另外从预测误差图中也抽取相应特征以提高性能,然后采用 SVM 支持向量机作为分类器进行图像分类,最后提取待检测图像和预测图像经验矩的 PDF 矩和 CF 矩共 108 维特征。实验结果表明:采用的 108 维特征分类效果很好,对隐写图像的平均正确分类率达 98% 以上。

9. 基于马尔可夫模型的通用隐写分析方法

Sullivan 等^[116]利用马尔可夫 (Markov) 链研究图像的空间相关性,提出了一种新的隐写分析方法,该方法根据图像邻域相关的性质构造马尔可夫链,以灰度共生矩阵主对角线上及其附近的元素作为特征。该方法可以有效地检测基于扩频的隐写算法,但是对于数据嵌入率较低的 QIM 以及 LSB 等隐写算法,其检测效果下降许多。孙子文等人于 2009 年在控制与决策第 24 卷第 8 期提出一种针对 JPEG 图像隐写的通用隐写分析方法。该方法根据量化后分块 DCT 系数绝对值构造垂直、水平和 zigzag 方向的差分数组,可利用三向差分数组马尔可夫模型挖掘量化后分块 DCT 块内邻近系数相关性,提取转移概率矩阵的特征。对三向特征加权融合后进行隐写分析,以提高分类性能。对于安全性较高的 JPEG 隐写方法 OutGuess 和 F5,在不同嵌入率下进行隐写分析,引入特征融合后隐写分析的检出率提高显著。Shi 等人^[117]运用马尔可夫链对频域系数的相关性建模,提出了基于 Markov 过程的 JPEG 图像的隐写分析方法。由于符号未参与隐写操作,为减少计算的复杂性,对频域块 DCT 系数取绝对值,定义了 JPEG 二维数组。通过对 8×8 块内的频域系数分别进行横向、纵向、主对角以及副对角方向的扫描构造了 4 个 Markov 链,分类特征由 4 个方向的 Markov 矩阵特征组成,并使用二类 SVM 作为分类器。尽管该算法对 Outguess 和 MB 获得了不错的检测效果,但是对低嵌入率的 F5 隐写算法检测效果不是太理想。

10. 基于多域联合特征的通用隐写分析方法

Lie 等^[118]提出了一种基于空域特征和 DCT 域特征联合通用检测方法,该方法从

空域中提取了梯度能量作为特征,从 DCT 域中提取图像中宏块的 Laplacian 参数的均值和方差为特征,将两域特征联合组成特征向量,用 3 层神经网络进行分类,从而对图像中是否含有隐藏信息作出判断。实验结果表明该方法对空域和 DCT 域隐藏方法有较好的检测效果,但是对小波域上的隐藏,该方法无效。且对原始图像的正确分类率相对较低,仅 70%左右。为了扩大 Lie 方法的适用范围,除了利用空域特征和 DCT 域特征以外,还可引入 DWT 域特征,Luo 等^[119]在 Lie 等^[118]方法的基础上,对空域特征提取方法进行了改进,并结合 DWT 域特征,采用 BP 神经网络分类器,给出了一种基于三域特征的图像信息隐藏通用检测方法。该方法在空域中使用 SPA 方法提取图像偏离度为特征,DCT 域提取特征方法和 Lie 等^[118]方法一样,在小波域中首先对图像进行 Haar 小波分解,并从分解得到的 6 个高频自带系数中提取前四阶矩(均值、方差、偏斜度和峰度)作为特征。实验结果表明三域联合特征方法对原始图像的正确检测率优于 Lie 等^[118]的方法,对隐密图像和原始图像的判断率更趋于平衡。

11. 基于纹理分类方法的通用隐写分析方法

Lafferty 等^[120]提出利用纹理分类的方法进行隐写分析。他们利用一种在纹理分析中常用的局部二值模式(Local Binary Pattern, LBP)对图像进行分析,提取 LBP 直方图作为隐写分析特征。由于文章中对该方法的介绍不够详细,且此方法仅对彩色图像效果明显,故其未被广泛采用。毛家发等通过提取两种新颖特征:纹理特征和虚特征值分解(IED)特征,对净图进行完整的定量描述,并且设计了与特征相匹配的柔性化超椭球体一类分类器,提高了检测性能。实验结果表明:作者方法对大多数含密图像和干净图像的检测效果都较好,但是对经过再压缩处理和锐化的图像,检测效果很不理想,虚警率过高。为了提高图像信息隐藏正确检测率,扩展隐写分析算法的适用范围,陈光喜等于 2009 年在清华大学学报自然科学版第 49 卷第 8 期提出了一种基于最低有效位(LSB)的隐写分析方法。该方法引入了一组基于相邻像素相关性和图像纹理复杂度差值关系的高阶统计矩作为特征矢量,采用支持向量机(SVM)进行训练和分类。通过仿真实验例证,针对原始无损存储图像,其方法建立的分类器的准确率高于目前的主流算法并具有较可靠的检测性能。

7.4.4 典型通用隐写分析算法

由于 JPEG 图像格式图像在互联网中广泛使用,而频域隐写算法通常可以在完成较大信息嵌入的前提下拥有较高的安全性,使得对 JPEG 图像隐秘信息的检测成为了通用隐写分析中近年来的研究热点。上一小节已经对目前现有的一些通用隐写分析算法做了概述,这一小节将重点介绍两种典型的 JPEG 图像通用隐写分析算法。

1. 基于 JPEG 兼容性的隐写分析

由 Fridrich 等提出的基于 JPEG 兼容性的隐写分析算法^[121],是通过研究已知量化表的 JPEG 图像的 8×8 块的兼容性,来实现对图像隐写前后变换的检测。

设 JPEG 图像的量化 DCT 系数集合为 $D=\{D_1, D_2, \dots, D_T\}$, T 为 DCT 系数块的总数目, $D_k=\{D_k(1), D_k(2), \dots, D_k(64)\}$ ($k=1, 2, \dots, T$) 为第 k 个量化 DCT 系数块。 D 逆量化后的 DCT 系数集合为 $QD=\{QD_1, QD_2, \dots, QD_T\}$, $QD_k(i)=D_k(i)Q(i)$, $Q(i)$ 为量化表 Q 的第 i 个系数, $i=1, 2, \dots, 64$ 。对 QD 作 DCT 逆变换得到的集合记作 $B_{\text{raw}}=\text{DCT}^{-1}(QD)$, 最终输出的解压图像记作 $B=[B_{\text{raw}}]$, 其中 $[]$ 表示取整到 $0 \sim 255$ 范围, 即当 $B_{\text{raw}, k(i)} \leq 0$, $B_k(i)=0$; $B_{\text{raw}, k(i)} \geq 255$, $B_k(i)=255$; 当 $0 < B_{\text{raw}, k(i)} < 255$, $B_k(i)$ 为 $B_{\text{raw}, k(i)}$ 取整值。定义

$QD' = DCT(B)$ 。由于取整的关系, 对于不包含饱和元素 (0 或 255) 的像素块 (称为非饱和块), $B_k(i)$ 和 $B_{raw,k}(i)$ 的误差不会超过 0.5, 从而 $\|B_k - B_{raw,k}\| \leq 64 \times 0.5^2 = 16$ 。于是, 对于非饱和块, 由于频域能量等于时域能量 (Parseval 等式), 可得

$$\|B_k - B_{raw,k}\|^2 = \|DCT(B_k) - DCT(B_{raw,k})\|^2 = \|QD'_k - QD_k\|^2 \leq 16 \quad (7.83)$$

另一方面, 通过将 $QD_k(i)$ 用 B_k 的第 i 个 DCT 系数的量化值乘以 $Q(i)$ 替换, 可以对 $\|QD'_k - QD_k\|^2$ 的下界进行如下估计

$$16 \geq \|QD'_k - QD_k\|^2 \geq \sum_{i=1}^{64} \left| QD'_k(i) - Q(i) \text{round} \left(\frac{QD'_k(i)}{Q(i)} \right) \right|^2 = S_k \quad (7.84)$$

这样, 在量化矩阵 Q 已知的情况下, 根据 B_k , S 值可用式 (7.84) 计算得出。

若 $S_k > 16$, 则说明非饱和块 B_k 与基于量化矩阵 Q 的 JPEG 格式不相容, 可判定秘密信息存在; 若 $S_k \leq 16$, 对非饱和块 B_k 还需进行进一步判断。设 $q_{k,l}(i)$, $l=1, 2, \dots$ 为离 $QD'_k(i)$ 最近的整数倍 $Q(i)$ 的序列并按照离 $QD'_k(i)$ 的距离远近进行排序 ($q_{k,1}(i)$ 离得最近), 用 $p(i) \in \{1, 2, \dots\}$ 来记录对应的一个可能索引。为了判断给定的非饱和图像块 B_k 是否与基于量化表 Q 的 JPEG 压缩兼容, 需要考虑所有满足下式的 64 元素索引组 $\{p(1), p(2), \dots, p(64)\}$

$$\hat{S}_k = \sum_{i=1}^{64} |QD'_k(i) - q_{k,p(i)}(i)| \leq 16 \quad (7.85)$$

令 $QD_k(i) = q_{k,p(i)}(i)$, $i=1, 2, \dots, 64$ 时, 检查是否满足下式

$$B_k = [DCT^{-1}(QD_k)] \quad (7.86)$$

若至少有一个 $\{p(1), p(2), \dots, p(64)\}$ 索引组满足上式, 则块 B_k 就是 JPEG 兼容的, 否则就不兼容。满足式 (7.85) 的 64 元素索引组的数目是有限的, 但是它们随着 JPEG 质量因子的提高而快速增加。对于大于 95 的质量因子, 大多数量化系数 $Q(i)$ 变为 1 和 2, 这时 64 元素索引组的数量将太多而难以处理。

基于 JPEG 兼容性的隐写分析算法流程可描述如下。

① 把图像分割成由 8×8 图像块组成的网格。若行数或列数不是 8 的整数倍, 则去掉多出的那几行或那几列。

② 把图像块排成一个序列, 从序列中排除所有饱和的图像块 (至少包含一个或多个灰度值为 0 或 255 的像素的块), 设序列中图像块的数量为 T 。

③ 从所有 T 个图像块中抽取出所使用的量化矩阵 Q , 如果 Q 中所有的元素都是 1, 则图像之前没有经过 JPEG 压缩, 隐写检测算法不适用而退出。如果存在一个或多个量化矩阵可供选择, 对每一个合理的量化矩阵执行步骤④到⑥, 其中最大的 JPEG 兼容图像块的数目将作为算法的输出结果。

④ 对于每一个图像块 B_k 根据公式 (7.84) 计算 S_k 。

⑤ 如果 $S_k > 16$, 则图像块 B_k 与基于量化矩阵 Q 的 JPEG 压缩并不兼容, 如果 $S_k \leq 16$, 则对每一个 DCT 系数 $QD'_k(i)$, 计算离 $QD'_k(i)$ 最近的整数倍 $Q(i)$ 的序列 $q_{k,p(i)}(i)$ 。对于每一个索引组合 $\{p(1), p(2), \dots, p(64)\}$, 检查当等式 (7.85) 成立时, 等式 (7.86) 是否成立。如果至少有一个索引组合使等式成立, 则图像块 B_k 是 JPEG 压缩兼容的, 否则不是。

⑥ 检查完所有 T 个图像块后, 如果没有找到非兼容的 JPEG 图像块, 则隐写分析算法并没有找到秘密信息嵌入的证据。否则, 如果找到一些兼容的 JPEG 图像块, 则我们可以试图估计秘密信息的长度, 定位存在秘密信息的像素, 甚至试图得到潜入秘密信息前的原始载体图像。

⑦ 若所有图像块都被识别为 JPEG 不兼容块, 或者图像似乎预先没有经过 JPEG 压缩, 则应该对图像进行不同的 8×8 分割 (从 X 和 Y 方向分别进行 0 至 7 个像素的平移) 来重复上述算法。这一步主要针对那些在嵌入秘密信息前进行过剪切操作的载体图像。

2. 基于小波特征函数统计矩的隐写分析

基于小波特征函数统计矩的隐写分析法^[111], 是一种通用型的隐写分析方法, 它使用小波子带的特征函数的统计矩作为隐写分析的特征。定义一幅图像的直方图 $h(x_l)$, $l=1, 2, \dots, N-1$ 对应的离散傅里叶变换为

$$H(f_k) = \sum_{n=0}^{N-1} h(x_l) e^{-j \frac{2\pi}{N} lk} \quad (7.87)$$

$H(f_k)$ 称为特征函数, $k=0, 1, \dots, N-1$, N 为图像的灰度数目。 $H(f_k)$ 的 n 阶统计矩定义为

$$M_n = \frac{\sum_{k=0}^{N/2} f_k^n |H(f_k)|}{\sum_{k=0}^{N/2} |H(f_k)|} \quad (7.88)$$

其中 $|H(f_k)|$ 是特征函数第 k 个频率成分的幅值。根据离散形式的 Chebyshev 不等式, 可以证明用隐写技术嵌入秘密信息后, M_n 的值将下降。

在表 7.2 中, 针对连续意义上的直方图, 对直接根据图像直方图计算的前 3 阶矩与根据图像直方图的特征函数计算的前 3 阶矩做了一个比较。具体来说, 基于特征函数的 n 阶矩正比于 $1/\sigma^n$, 而基于直方图的 n 阶矩正比于 σ^n 。因此, 基于特征函数的矩对隐密信息的标准方差的变化更为敏感。

表 7.2 基于直方图的矩与基于特征函数的矩的比较

n 阶绝对矩	$n=1$	$n=2$	$n=3$
原始直方图域 (高斯分布) $\int_{-\infty}^{+\infty} x ^n h(x) dx, \quad h(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$	$\frac{\sqrt{2}\sigma}{\sqrt{\pi}}$	σ^2	$\frac{2\sqrt{2}\sigma^3}{\sqrt{\pi}}$
特征函数域 (高斯分布特征函数) $\frac{\int_{-\infty}^{+\infty} f ^n \cdot H(f) df}{\int_{-\infty}^{+\infty} H(f) df}, \quad H(f) = e^{-\frac{\sigma^2 f^2}{2}}$	$\frac{\sqrt{2}}{\sqrt{\pi}\sigma}$	$\frac{1}{\sigma^2}$	$\frac{2\sqrt{2}}{\sqrt{\pi}\sigma^3}$

基于上述考虑, 基于小波特征函数统计矩的隐写分析法对待检图像作 Haar 离散小波变换, 以待检图像本身及每一个子带的前 3 阶矩作为统计特征, 构建一个 39 维的特征向量, 使用贝叶斯 (Bayes) 分类器进行分类。算法的过程可以描述如下。

首先对待检测图像进行三层 Haar 小波变换, 包括图像本身在内 (这里看做子带 LL0), 将获得 13 个子带: LL0、LL1、HL1、LH1、HH1、LL2、HL2、LH2、HH2、LL3、HL3、LH3、HH3。然后根据式 (7.87) 计算每个子带对应直方图的 DFT, 即得到了特征函数。最后根据式 (7.88), n 取 1、2、3, 可以得到一阶、二阶和三阶矩, 这样就构建了一个 39 维的特征向量。

Bayes 分类器的分类原理是通过某对象的先验概率, 利用贝叶斯公式计算出其后验概率, 即该对象属于某一类的概率, 选择具有最大后验概率的类作为该对象所属的类。由于嵌入的秘密信息一般服从高斯分布或近似于高斯分布, 因此选用 Bayes 分类器。

假设用 X_i 表示第 i 幅图像的特征向量, I_1 与 I_2 分别表示原始图像集和隐写图像

集, 其均值向量和协方差矩阵分别由 μ_1 、 μ_2 和 Σ_1 、 Σ_2 表示, 则 Bayes 分类描述如下。

(1) 最大后验决策

如果 $P(I_1|X_i) \geq P(I_2|X_i)$, 则 $X_i \in I_1$, 否则 $X_i \in I_2$, 其中

$$P(I_k | X_i) = \frac{P(I_k)P(X_i | I_k)}{\sum_{m=1}^2 P(I_m)P(X_i | I_m)}, \quad k=1,2 \quad (7.89)$$

并且

$$P(X_i | I_k) \propto N(X_i, \mu_k, \Sigma_k), \quad k=1,2 \quad (7.90)$$

其中 $N(X_i, \mu_k, \Sigma_k)$ 表示均值为 μ_k 和方差为 Σ_k 的正态分布。

(2) 决策函数

若 $g_1(X_i) \geq g_2(X_i)$, 则 $X_i \in I_1$, 否则 $X_i \in I_2$, 其中

$$g_k(X_i) = -0.5X_i^T \Sigma_k^{-1} X_i + \left(\Sigma_k^{-1} \mu_k \right)^T X_i - 0.5\mu_k^T \Sigma_k^{-1} \mu_k - 0.5 \ln |\Sigma_k|, \quad k=1,2 \quad (7.91)$$

大量实验证明, 基于小波特征函数统计矩的隐写分析方法很有效, 检测率很高。下面, 给出一个例子来说明小波特征函数统计矩的有效性。图 7.11 (a) 是 CorelDraw 图像数据库中的一幅彩色图像 (第 18093 号), 图 7.11 (b) 是对应的灰度图像, 图 7.11 (c) 是采用 Cox 的扩频方法嵌入秘密信息后的效果图。图 7.12 显示了该灰度图像嵌入秘密信息前后的一层小波变换的四个边带的直方图, 图 7.12 (a) 为整体直方图, 图 7.12 (b) 为直方图的局部放大。图 7.13 给出了对应的特征函数。由图 7.12 和图 7.13 可以发现, 隐写图像的小波边带直方图比原始图像的直方图要平坦。通过观察图像一阶矩, 明显发现隐写图像的一阶矩要小于原始图像的一阶矩。也就是说, 在数据隐藏过程之后, 隐写图像直方图在 $x=0$ 处的一阶导数的最大值要低于原始图像在该点导数的最大值, 这一点也可以从图 7.13 中看出。

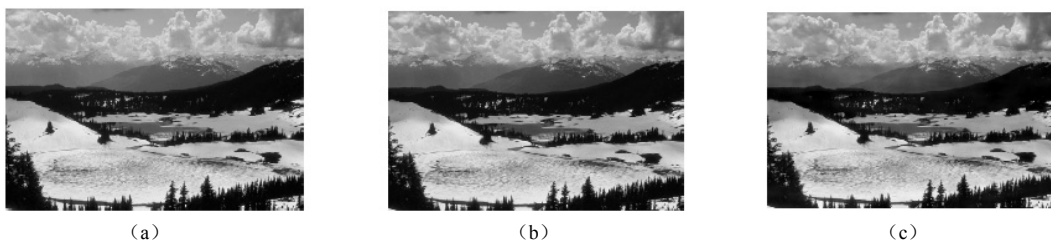


图 7.11 CorelDraw 图像库第 18093 号图像隐写效果

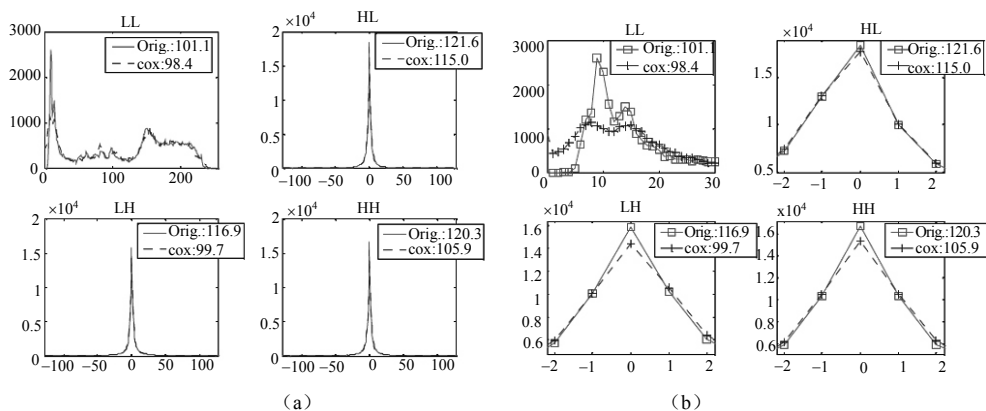


图 7.12 CorelDraw 图像库第 18093 号图像嵌入秘密信息前后的一层小波边带的直方图比较

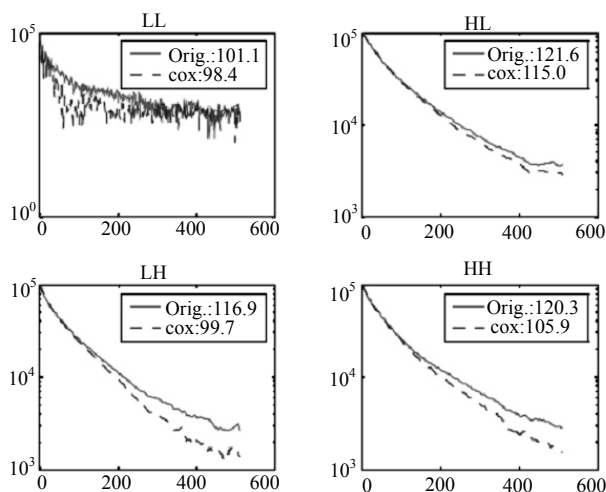


图 7.13 CorelDraw 图像库第 18093 号图像嵌入秘密信息前后的特征函数比较

7.5 音频隐写分析技术

在 2.6 节介绍了一些常用的音频隐写方法和隐写工具。对音频的隐写会对音频信号某些方面的特征产生影响，而隐写分析正是利用这种影响来区分原始音频和隐写音频的。本节首先分析不同隐写方法对音频的影响以及隐写对压缩音频的影响，这是专用隐写分析的基础；然后分析了隐写对音频的综合影响，这是通用隐写分析的基础；接着介绍音频隐写分析系统的通用框架模型；然后介绍几种针对音频 LSB 嵌入的专用隐写分析方法；最后介绍几种典型的音频通用隐写分析方法。

7.5.1 专用音频隐写分析的机理

由于 2.6 节中介绍的音频隐写方法较多，而 LSB、ECHO、DSSS 和 Mp3Stego 这几个专用隐写方法比较有代表性，所以本节只选取它们来进行分析。隐写对音频的影响是多方面的，这里只讨论本书所涉及的隐写方法对音频产生的特定的影响，可以作为专用隐写分析的依据。

1. LSB 隐写对音频信号的影响

在 LSB 隐写中，音频时域值有可能由 $2i$ 改为 $2i+1$ 或由 $2i+1$ 改为 $2i$ ，而不会将 $2i$ 改成 $2i-1$ 或将 $2i+1$ 改为 $2i+2$ ，所以隐写后音频信号的直方图会发生改变。设原始音频中，时域信号值为 i 的数目为 f_i ，则 LSB 隐写会减少 f_{2i} 与 f_{2i+1} 之间的差距，而在 f_{2i} 与 f_{2i-1} 之间不会存在这种情况。如果音频载体没有被隐写， f_{2i} 与 f_{2i+1} 之间会相差的远一些，而经过隐写的音频， f_{2i} 与 f_{2i+1} 的差距较小。 χ^2 分析法就是根据这个特点进行隐写分析的。

LSB 隐写还会对音频的差分直方图产生影响。由于 LSB 隐写只是替换了音频信号的最低位，原始音频和隐写音频的外形轮廓并没有太大的差别。而将原始音频和隐写音频的 LSB 平面置反后，两者之间的差分直方图产生明显的区别，原始音频在 LSB 平面置反后的差分直方图发生明显的变化，而隐写音频在 LSB 平面置反后的差分直方图几乎没有改变。基于差分直方图的方法 (DIH) 就是根据这个差异进行隐写分析的。

音频信号的位平面之间存在着一定的相关性，LSB 隐写会破坏这种相关性。定义两

种置换操作： F_1 为 $2i$ 与 $2i+1$ 的转换关系， F_{-1} 为 $2i-1$ 与 $2i$ 的转换关系。对原始音频无论应用 F_1 变换还是 F_{-1} 变换，都会同等程度的破坏这种相关性。而对于 LSB 隐写音频，应用 F_1 变换和 F_{-1} 变换之后的结果有明显的差别，应用 F_1 变换部分样本回到了原来的值，而应用 F_{-1} 样本只会更偏离原来的值，进一步破坏相关性。RS 分析法就是根据这个特性进行隐写分析的。

在自然音频中，不同抽样对组成的多个集合的势有相同的期望值。例如： $P=\{(s_i, s_j) | 1 \leq i, j \leq N\}$ 表示音频信号的抽样对集合。抽样对集合 $X=\{(u, v) | (u, v) \in P, u < v \text{ 且 } v \text{ 为偶数或 } u > v \text{ 且 } v \text{ 为奇数}\}$ 与 $Y=\{(u, v) | (u, v) \in P, u > v \text{ 且 } v \text{ 为偶数或 } u < v \text{ 且 } v \text{ 为奇数}\}$ 的势有相同的期望值，即 $E\{|X|\}=E\{|Y|\}$ 。LSB 隐写会使得抽样对在这些集合间转换，从而改变了这些集合的势。这个特性构成了**抽样对分析法**（Sample Pairs Analysis, SPA）的算法基础，通过将集合势的改变拟合成一个关于嵌入率的二次方程，从而能够估计出隐写的嵌入率。

在下面的 7.5.3 节中，将给出几种经典的 LSB 隐写分析方法： χ^2 分析法、RS 分析法和 SPA 分析法，根据 LSB 隐写机理引出了具体的算法实现。这几种隐写分析方法不仅能够判断出待测媒体是否含有秘密信息，还能估计出秘密信息的长度。

2. ECHO 隐写对音频信号的影响

回声隐写与 LSB 隐写不同，它不是以随机噪声形式将秘密信息嵌到音频载体中，而是利用音频的回声信号来嵌入秘密信息的。回声信号可以看作音频信号的衰减，并滞后于音频信号。回声隐写将秘密信息调制成回声信号，叠加到音频信号上，即回声隐写的算法可表示为 $x(n)=s(n)+\alpha s(n-d)=s(n)*h(n)$ ，其中 $h(n)=\delta(n)+\alpha\delta(n-d)$ 为回声核。由于隐写后的音频信号为原始音频信号与回声核的卷积，对隐写后的音频信号求复倒谱可得到

$$c_x(n) = c_s(n) + \alpha\delta(n-d) - \frac{1}{2}\alpha^2\delta(n-2d) + \frac{1}{3}\alpha^3\delta(n-3d) - \dots \quad (7.92)$$

可以看出，回声隐写方法对于音频信号的影响在复倒谱系数中有明显的体现，即：加入回声后音频的复倒谱会在回声延迟的整数倍产生峰值，可以根据这个特点进行隐写分析。由于复倒谱计算较为复杂，还可以利用功率倒谱来检测回声。与复倒谱相比，功率倒谱的计算相对简单，且在回声延迟处所产生的峰值更为明显。

3. DSSS 隐写对音频信号的影响

DSSS 隐写方法用 PN 序列来调制秘密信息序列，得到扩展频谱序列，再乘以一定的衰减系数，叠加到每一帧的载体音频信号上去。扩频隐写方法将秘密信息的能量扩展到载体信号的整个频率上，而分配到每个频率分量上的能量都很小，增加了秘密信息的隐蔽性。

DSSS 隐写对音频信号的一个显著影响体现在频域上。一般情况下，原始音频信号的能量主要集中在中低频部分，而高频部分的能量很小，有时可以忽略，音频压缩通常利用这个性质，将高频部分的系数置为零，减少编码的信息量，从而达到音频压缩的目的。由于扩频隐写将秘密信息的能量扩展到音频信号的整个频率上，音频信号中低频部分本身的能量就较大，隐写所增加的能量相对不明显，而高频部分本身的能量很小，因此经过扩频隐写后其能量明显增加了，可以根据这个特点来进行隐写分析。例如，根据隐写对高频部分的影响比低频部分显著的特点，可采用小波包的高频系数作为分析对象，来进一步体现出原始音频信号与隐写音频信号在高频部分的差别。或者，利用短时傅里叶变换，先将音频信号取二阶差分，然后再变换到频域中，可减少隐写对低频部分的影响，而扩大对高频部分的影响，使得隐写对高频部分的影响比低频部分显著这一特

点得到了进一步的强化, 增加所提取特征的分度。

DSSS 隐写对音频信号的另一个影响体现在 PN 序列的自相关性上。在提取秘密信息时, 接收方根据产生 PN 序列的密码种子, 可以重新生成该 PN 序列, 然后将该 PN 序列与收到的隐写音频做相关运算来提取秘密信息。由于 PN 序列只与其自身相关, 其自相关值为 1, 与其他的序列都不相关, 因此, 对于原始音频信号, 无论 PN 序列是怎样的, 其相关值均接近于 0。而对于 DSSS 隐写音频信号, 如果 PN 序列与扩频序列一致或相反, 长度相等时, PN 序列与隐写音频信号的相关值就会出现峰值。可以根据这一特点, 对 PN 序列及其长度进行估计, 通过观察是否出现峰值及峰值出现的位置可以判断待测载体是否经过 DSSS 隐写, 甚至进一步提取出秘密信息。

4. MP3Stego 隐写对压缩音频的影响

前面讨论了隐写对于未压缩音频的影响。随着网络与多媒体技术的发展, 以 MP3 和 AAC 为代表的感知编码技术得到了广泛的应用, 目前网络上传播的音频大多是压缩格式的音频。因此, 以压缩音频为载体的隐写方法具有实用意义, 开始受到有关研究人员的重视。目前压缩音频的隐写方法的文献较少, 最具代表性的是 MP3Stego, 它是在压缩过程中实现秘密信息嵌入的。这里以 MP3Stego 为例来分析隐写对压缩音频的影响。

MP3Stego 通过对量化模块的内循环结束条件进行修改实现秘密信息的嵌入, 根据经过量化编码得到的数据块长度 (part_2_3) 的奇偶性与待嵌入比特是否一致, 来决定是否继续量化。因此, 经过 MP3Stego 隐写后的 MP3 编码的 part_2_3 长度与没有经过隐写的 MP3 编码的 part_2_3 长度有所不同。在正常的 MP3 编码过程中, 当 part_2_3 长度小于最多可用比特数的要求后就可以结束内部循环, 但是在 MP3Stego 隐藏方式下, part_2_3 长度还需要满足秘密信息嵌入的要求, 因此可能需要进一步扩大量化步长, part_2_3 长度也会随之减少, 即这一帧编码所用的比特数减少。由于 MP3 编码的帧长度和帧速率是恒定的, 下一帧可用的比特数就增大, 因此, 下一帧 part_2_3 长度也会随之变大。经过 MP3Stego 隐写后, 虽然平均 part_2_3 长度保持不变, 但是 part_2_3 长度的方差发生了变化。根据这个机理, 可采用 part_2_3 长度的方差作为特征来进行隐写分析。

秘密信息的嵌入使得量化步长得到了改变, 因此, 经过 MP3Stego 隐写后的 MP3 音频的量化后 MDCT 系数也发生了改变, 我们还可以根据隐写后量化后 MDCT 系数的变化来进行隐写分析。一方面, 量化后 MDCT 系数的变化导致量化直方图的变化, 可以提取直方图的统计量来作为特征。另一方面, 隐写对量化 MDCT 系数的相关性产生了影响, 可以采用马尔可夫转移矩阵来衡量 MDCT 系数的相关性。

7.5.2 通用音频隐写分析的机理

在上一节中, 我们讨论了特定隐写方法对音频信号的影响, 根据这个机理, 可以设计专用隐写分析方法。本节讨论多种隐写方法对音频信号某些方面特征的影响, 这种影响不局限于某一种隐写方法, 可以据此来设计通用隐写分析方法。当然, 隐写对于音频的影响是多方面的, 这里不可能面面俱到, 只选取有代表性的特点进行讨论。

1. 隐写对音频相关性的影响

音频信号的时域相邻样本值之间存在着一定的相关性, 而且采样频率越高, 相邻样本值之间的相关性就越强。此外, 音频信号在频域上也存在着一定的相关性。而秘密信息的嵌入会改变这种相关性, 因此, 在隐写分析中, 根据音频信号的时域和频域存在着相

关性这一机理, 提取度量这种相关性的特征参数, 可以对原始音频和隐写音频进行区分。衡量这种相关性可以采用预测的方式。其中, 线性预测是在音频信号处理中是一类重要的处理技术, 它能挖掘信号前后样点之间的相关性。线性预测器使用过去的 p 个样本值来预测当前的样本值, 预测值可以用过去 p 个样本值的线性组合来表示。在自然音频中, 前后样点的相关性很强, 因此波形较为平滑, 而秘密信息的嵌入操作则会改变这种相关性, 使得波形的平滑度降低。这就导致原始音频的可预测性要优于隐写音频, 因此隐写音频的预测误差较原始音频大。例如, 我们可以采用自适应预测器来提取出音频信号的相关性, 对采用小波包分解的高频子带系数的自适应预测误差求统计特征。

此外, 隐写的次数对音频相关性也有影响。第一次隐写对这种相关性的破坏程度, 要大于其后所进行的隐写操作。因此, 如果待测的音频中已经过隐写, 再对它进行隐写的话, 对音频相关性的影响就会减少, 如果待测音频本身没有经过隐写, 那么对它进行隐写操作对相关性的破坏就较大。因此, 可以对待测音频再次隐写, 然后通过比较对相关性的破坏程度来进行隐写分析。

2. 隐写对音频直方图的影响

以加性噪声隐写方法为例, 由于在加性噪声模型中, 音频载体与秘密信息所产生的噪声是相互独立的, 因此音频隐写后的直方图相当于原始音频直方图与秘密信息产生的噪声直方图的卷积。隐写使得音频直方图发生了变化, 卷积效果在时域直方图上表现为: 隐写后, 音频直方图的尖峰变平, 直方图向两端扩展。可以根据这个特点, 利用直方图统计特征的变化来进行隐写分析。直方图的统计特征可以考虑采用直方图的多阶统计矩, 但是在空域直方图中, 统计矩反映直方图的总体变化情况, 如果隐藏信息较少时, 直方图的总体变化不大, 因此, 空域统计矩的变化较少, 不足以作为隐写分析的显著特征。

如果将直方图变换到频域上, 则时域的卷积就变成了频域的乘积。隐写后, 相当于对直方图进行了一次低通滤波, 而直方图的离散傅里叶变换可看作直方图的特征函数, 因此隐写后直方图特征函数质心前移, 即隐写后音频直方图特征函数的质心较隐写前的小。因此, 可以通过设置一个阈值并比较音频直方图特征函数的质心来判断是否隐写。此外, 还可以用直方图频域的多阶统计矩作为统计特征, 频域矩反映的是直方图尖峰处的变化情况, 即使隐藏的信息较少, 在直方图的尖峰处的变化还有比较明显的, 因此, 音频直方图的频域矩比时域矩的变化更为明显, 对于隐写更为敏感。前面介绍的基于直方图特征函数统计矩的隐写分析方法就可以推广到音频中, 根据隐写引起音频直方图的变化这一机理来设计具体隐写方法。

3. 隐写对音频质量测度的影响

隐写将秘密信息以一定的方式嵌入到载体音频中, 对原始音频进行了修改必然会引起音频质量的下降, 虽然人耳对这种音频质量的影响不敏感, 但是可以通过一系列音频质量测度来揭示秘密信息的存在性。利用小波去噪或者预测等方法对音频信号进行估计, 由于原始音频信号波形较为平滑, 原始音频信号与其估计信号之间的差异较小, 相当于音频质量的下降不明显, 因此, 它们之间的音频质量测度较小。而经过隐写后的音频平滑度降低, 估计信号与原来信号差异增大, 因此, 隐写后音频信号与其估计之间的音频质量测度增大。从而可知, 隐写音频信号与其估计之间的音频质量测度大于原始音频信号与其估计之间的音频质量测度, 可以通过选取一定的音频质量测度作为特征来进行隐写分析。可供选取的音频质量测度见表 7.3 所示, 可以分为感知域和非感知域两类, 感知域测度从人耳的听觉特性方面来考虑的, 而非感知域测度可以看作被测信号与

参考信号之间的某种距离。非感知域测度又可以分为时域测度和频域测度，分别在时域和频域两方面对音频质量的变化进行度量。下面给出表 7.3 中一些质量测度的定义。

(1) 加权频谱斜率距离 (Weighted Spectral Slope Distance, WSSD)

将音频信号的频带划分为 25 个 Bark 带，设 $X_n(k)$ 和 $Y_n(k)$ 分别表示原始信号 X 和失真信号 Y 的第 n 帧信号的第 k 个 Bark 带的频谱能量，则 $V_{x,n}(k)$ 和 $V_{y,n}(k)$ 分别表示 X 和 Y 的第 n 帧信号的第 k 个 Bark 带频谱能量斜率，即 $V_{x,n}(k) = X_n(k+1) - X_n(k)$ 和 $V_{y,n}(k) = Y_n(k+1) - Y_n(k)$ 。则 WSSD 定义如下

$$\text{WSSD} = \frac{1}{N} \sum_{n=1}^N \sum_{k=1}^{25} w(k) [V_{x,n}(k) - V_{y,n}(k)]^2 \quad (7.93)$$

其中， $w(k)$ 为第 k 个 Bark 带的权重。由于音频信号的频域能量在低频部分比高频部分大得多，为了使得 WSSD 更能够反映出高频部分的失真，故这里 $w(k)$ 设定为第 k 个 Bark 带能量的倒数，帧长取 1024。

表 7.3 音频质量测度列表

感知域测度	非感知域测度	
	时域测度	频域测度
Bark 域频谱失真 (BSD)	信噪比 (SNR)	对数似然比 (LLR)
修正的 BSD (MBSD)	分段信噪比 (SNRseg)	对数面积比 (LAR)
增强的 MBSD (EMBSD)	Czenakowski 距离 (CZD)	Itakura-Siata 距离 (ISD)
感知语音质量测度 (PSAQ)		COSH 距离 (COSH)
感知音频质量测度 (PSAQ)		倒谱距离 (CD)
标准化段测量 1 (MNB1)		短时 Fourier-Radon 变换距离 (STFRT)
标准化段测量 2 (MNB2)		频谱相位距离 (SPD)
加权频谱斜率距离 (WSSD)		频谱相位幅值距离 (SPMD)

(2) Bark 谱失真 (BSD) 和改进的 Bark 谱失真 (MBSD)

Bark 谱失真 (Bark Spectral Distortion, BSD) 定义为

$$\text{BSD} = \sum_{i=1}^K [S_x(i) - S_y(i)]^2 \quad (7.94)$$

其中， K 是临界频带数， $S_x(i)$ 和 $S_y(i)$ 分别是原信号与失真信号第 i 个临界频带 Bark 谱系数。

改进的 Bark 谱失真 (Modified BSD, MBSD) 根据掩蔽阈值得到了一个权值，然后用权值乘上 Bark 谱失真，即

$$\text{BSD} = \sum_{i=1}^K W(i) [S_x(i) - S_y(i)]^2 \quad (7.95)$$

其中 $W(i)$ 表示第 i 个 Bark 带根据掩蔽阈值得到的权值。

(3) 频谱相位失真 (SPD)

在频域上，频谱相位失真 (Spectral Phase Distortion, SPD) 定义为

$$\text{SPD} = \frac{1}{N} \sum_{n=1}^N |\theta_x(n) - \theta_y(n)|^2 \quad (7.96)$$

其中， $\theta_x(n)$ 和 $\theta_y(n)$ 分别表示原始信号 X 和失真信号 Y 的第 n 帧的相位向量，一共 N 帧，帧长通常取 1024。

(4) 频谱相位—幅度失真 (SPMD)

联合上面的 SPD, SPMD (Spectral Phase Magnitude Distortion) 定义为

$$\text{SPMD} = \frac{1}{N} \left(\lambda \sum_{n=1}^N |\theta_x(n) - \theta_y(n)|^2 + (1 - \lambda) \sum_{n=1}^N \|X(n) - Y(n)\|^2 \right) \quad (7.97)$$

其中, $X(n)$ 和 $Y(n)$ 分别为原始信号 X 和失真信号 Y 的第 n 帧的幅值向量, $\theta_x(n)$ 和 $\theta_y(n)$ 同样为 X 和 Y 的第 n 帧的相位向量, $\lambda \in [0, 1]$ 为权重, 表示 SPMD 取决于幅值失真多一些还是相位失真多一些。

(5) 对数似然比 (LLR)

LLR (Log Likelihood Ratio) 是根据音频段的线性预测模型得到的, 定义为

$$\text{LLR} = \log \frac{\mathbf{a}_x^T \mathbf{R}_x \mathbf{a}_x}{\mathbf{a}_y^T \mathbf{R}_y \mathbf{a}_y} \quad (7.98)$$

其中 \mathbf{R}_x 和 \mathbf{R}_y 分别为原始信号 X 和失真信号 Y 各自的协方差矩阵, \mathbf{a}_x 和 \mathbf{a}_y 分别表示 X 和 Y 的线性预测系数 (LPC) 向量。

(6) Itakura-Siata 距离 (ISD)

Itakura-Siata 距离定义为

$$\text{ISD} = \int_{-\pi}^{\pi} \left[\frac{1}{2} \left(\frac{Y(\omega)}{X(\omega)} + \frac{X(\omega)}{Y(\omega)} \right) - 1 \right] \frac{d\omega}{2\pi} \quad (7.99)$$

其中, $X(\omega)$ 和 $Y(\omega)$ 分别为原始信号 X 和失真信号 Y 的能量谱。采用 LPC (Linear Predictive Coding) 预测系数, ISD 可以表示为

$$\text{ISD} = \frac{1}{\text{LR}} + \text{LLR} - 1 \quad (7.100)$$

其中

$$\text{LR} = \frac{g_x}{g_y} \quad (7.101)$$

其中, g_x 和 g_y 分别为 X 和 Y 的 LPC 预测残差能量。

(7) 倒谱距离测度 (Cepstrum Distance Measure, CDM)

原始信号 X 和失真信号 Y 的第 n 帧信号的 LPC 系数倒谱距离定义为

$$d(\mathbf{c}_x, \mathbf{c}_y, n) = \sqrt{[c_x(0) - c_y(0)]^2 + 2 \sum_{k=1}^L [c_x(k) - c_y(k)]^2} \quad (7.102)$$

其中 L 为 LPC 系数倒谱的个数, \mathbf{c}_x 和 \mathbf{c}_y 分别为 X 与 Y 的第 n 帧信号的 LPC 预测系数倒谱向量。根据上式得到所有 N 帧的 $d(\mathbf{c}_x, \mathbf{c}_y, n)$, 则 CDM 的定义如下

$$\text{CDM} = \frac{\sum_{n=1}^N w(n) d(\mathbf{c}_x, \mathbf{c}_y, n)}{\sum_{n=1}^N w(n)} \quad (7.103)$$

其中, $w(n)$ 为第 n 帧权重。通常, 帧长为 1024 点, 系数倒谱长度 $L=10$, 预测阶数取 10, 采用当前帧 n 的能量作为 CDM 计算中每帧的权重 $w(n)$ 。

(8) 短时 (分段) 信噪比 (SNRseg)

SNRseg 定义为短时音频段信噪比的平均值

$$\text{SNR seg} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \sum_{i=Nm}^{Nm+N-1} \left(\frac{x^2(i)}{[x(i) - y(i)]^2} \right) \quad (7.104)$$

其中, M 为帧数, N 为帧长, $x(i)$ 为原始音频信号, $y(i)$ 为失真的音频信号。分段的长度一般为 15~20ms。SNRseg 用于能量高于特定阈值的音频段以避免静音段。

基于音频质量测度的通用音频隐写分析方法就是根据隐写对音频质量产生影响, 隐写音频与其估计之间的测度大于原始音频与其估计之间的测度这一机理, 利用特征选择方法选择对隐写敏感的特征, 从而对待测音频进行检测。

7.5.3 音频隐写分析系统模型

通用的隐写过程可表示为: $S=C+f(C, M)$, 式中 C 表示原始载体信号, S 代表嵌入秘密信息后的隐写信号, M 为待嵌入的秘密信息, 将待嵌入的秘密信息以某种方法嵌入到原始载体信号中, $f(C, M)$ 即为秘密信息的嵌入方法, 它既可以与原始载体信号无关 (如直接扩频隐写), 也可以与原始载体信号有关 (如回声隐藏)。隐写分析的过程就是从 S 中检测出是否含有秘密信息 M 甚至提取出 M 。隐写分析的通用系统模型如图 7.14 所示。

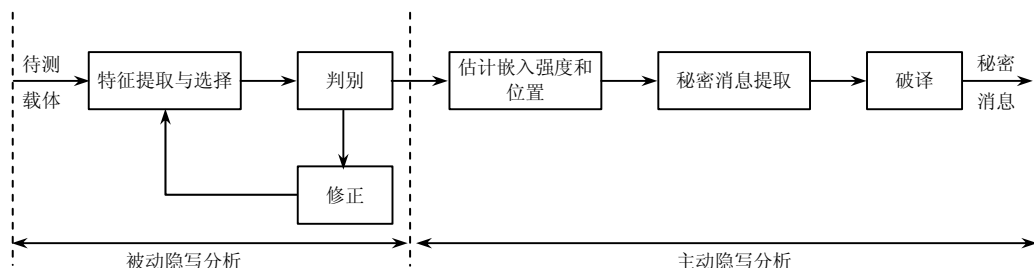


图 7.14 隐写分析系统通用模型

前面针对隐写分析技术分类, 曾提到过: 根据隐写分析实现的功能, 隐写分析分为被动隐写分析与主动隐写分析两个方面, 其中被动隐写分析是主动隐写分析的前提与基础。被动隐写分析只需检测被测载体中秘密信息的有无, 一般包括特征提取、特征选择和判别等环节。而主动隐写分析则是在检测到秘密信息存在的基础上估计秘密信息的嵌入强度和位置, 进一步对秘密信息进行提取和破译。主动隐写分析在隐写方法和密钥未知的情况下提取秘密信息当前还难以解决, 至今很少有相关的文献提及这方面的研究, 因此, 目前对于隐写分析的研究工作主要集中在被动隐写分析上。

1. 特征提取与特征选择

特征提取是隐写分析中的重要环节, 对待测载体信号进行特征提取, 通过对提取的特征进行分类来判别载体中是否含有秘密信息。特征提取包括特征的寻找与选择。特征提取是隐写分析的关键部分, 特征提取的好坏对隐写分析结果具有至关重要的影响。用于隐写分析的特征需要满足两方面要求: 一是隐写后与隐写前的特征有很好的区分度, 即所提取的特征对隐写过程有较高的敏感度, 二是所提取的特征与载体信号的内容有一定的独立性, 也就是说, 所提取的特征受载体内容的影响不大。这是因为载体内容是千变万化的, 而隐写分析的目的是检测出待测载体中是否含有秘密信息, 与载体本身没有关系, 如果特征受载体内容的影响较大, 隐写分析可能对某些特定的载体检测结果较好, 而对其他的载体检测结果不好, 这并不符合隐写分析的要求。因此, 在特征提取中, 需减少载体本身的影响。

特征选择就是对所提取的特征进行选择,以消除某些对隐写不敏感的特征的干扰,提高判别准确性。特征选择的好坏对分类器的性能有较大的影响,通常要求不同模式的类的特征值之间差别越大越好,而同一模式类中的特征值要接近。选择的各个特征之间要有独立性,如果特征之间有较大的相关性,则说明所选取的特征存在着冗余,这样一方面造成特征向量的维数过大影响分类器的效率,另一方面过分强调某一方面的特性而影响判别的准确率。常用的特征选择方法有方差分析法(ANalysis Of VAriance, ANOVA)、SFS(Sequential Floating Search)法、F-Score、PSO(Partical Swam Optimization, 粒子群优化算法)和主成分分析法(Principal Component Analysis, PCA)等。其中方差分析法是统计学中常用的数据处理方法,其目的是通过数据分析找出对该事物有显著影响的因素。ANOVA 的前提假设是:各样本是相互独立的随机样本,各样本均服从正态分布且方差相等。采用 ANOVA 法可以去除对隐写不敏感的特征,提高分类器的效率和判别准确率。PCA 是一种掌握事物主要矛盾的统计分析方法,其目的是将高维数据投影到较低维空间,并尽可能保留数据的信息。采用 PCA 法可以减少特征向量的维数,去除特征之间的相关性,提高分类的效率和准确性。

2. 判别

判别是根据提取的特征对待测载体信号进行归类、判断。特征提取的好坏、分类器的选择都将影响判别结果的好坏。在隐写分析技术中,模式识别和机器学习的理论得到了应用。隐写分析可分为学习过程和决策过程,学习过程是通过对大量学习样本所提取的特征进行比较和分析,建立起分类模型,决策过程是根据分类模型对待测样本进行检测和判断。在音频隐写分析中,首先将载体音频和对应的隐写音频组成学习样本库,然后从两类音频中抽取特征向量并训练分类器,最后使用同样的算法从测试音频库中抽取特征向量,根据训练好的分类器进行分类。通常采用支持向量机进行分类,也可采用神经网络、贝叶斯分类器、线性回归模型、Fisher 线性判别(FLD)等方法进行分类。修正是为了提高判别的准确性而采用的一个反馈过程,根据检测结果对所提取的特征、分类器所用的系数或阈值等做相应的改动,最终达到提高判别的准确性的目的。通常,隐写分析方法多采用支持向量机(SVM)作为分类工具。

7.5.4 针对 LSB 隐藏的隐写分析方法

隐写分析是一项比较困难的工作,因为一般情况下隐写分析者只能获得隐写文件,而对载体、嵌入方法、嵌入强度、嵌入位置、秘密信息、加密密钥等信息一无所知。因此,现阶段许多研究工作的主要目标还只能是检测出待测载体中是否有秘密信息,如果能够估计出嵌入秘密数据的长度,就已经是比较高的水平。要想成功把秘密数据提取出来,除了已知特定的隐写方法和密钥外,其他的还未见先例,目前的研究工作还主要集中在被动隐写分析上。虽然音频 LSB 隐写可以在隐藏大量信息的情况下依然保持良好的听觉隐蔽性,但是根据 LSB 隐写的特点,采用有效的统计分析工具就可以判断出一段音频是否含有秘密信息。以下介绍几种针对 LSB 隐写的有效分析方法。

1. χ^2 检测法

χ^2 检测法适用于嵌入信息均匀分布且连续嵌入的 LSB 隐写方法。设一段待测音频样本值为 i 的出现频率为 h_i , 其中 $i \in [0, 255]$ (采用 8 位 PCM 编码) 或 $i \in [0, 65535]$ (采用 16 位 PCM 编码)。根据 LSB 隐写算法的原理,嵌入将样本值由 $2i$ 改变为 $2i+1$, 或者

由 $2i+1$ 改变为 $2i$ ，而不会将 $2i$ 改变成 $2i-1$ 或将 $2i+1$ 改成 $2i+2$ 。LSB 隐写导致样本值变化的不规则性，使得音频直方图发生了改变，样本值或者不变，或者在 h_{2i} 和 h_{2i+1} 之间互变，所以 $h_{2i}+h_{2i+1}$ 的值是不变的。在嵌入之前秘密信息一般经过加密，成为 0、1 均匀分布的比特流。由于加密后的秘密信息位为 0 或 1 的可能性都是 1/2，如果所有载体信号的最低位都用秘密信息位代替，那么 h_{2i} 和 h_{2i+1} 的值会比较接近；而如果载体音频未经隐写，样本值的最低位不是完全的随机分布，因此 h_{2i} 和 h_{2i+1} 的值会相差得远一些。由上述可知，我们可以利用 h_{2i} 和 h_{2i+1} 值的分布来判断秘密信息的有无。

令 $h_{2i}^* = (h_{2i} + h_{2i+1})/2$ ，因为 LSB 隐写方法样本值只在 $2i$ 和 $2i+1$ 之间改变，所以 LSB 嵌入前后 h_{2i}^* 的值不变，对满容量嵌入的那部分载体进行分析，值为 $2i$ 的出现频率 h_{2i} 的期望值应该等于 h_{2i}^* 。当 h_{2i}^* 较大时，根据中心极限定理有

$$\frac{h_{2i} - h_{2i+1}}{2\sqrt{h_{2i}^*}} = \frac{h_{2i} - h_{2i}^*}{2\sqrt{h_{2i}^*}} \sim N(0,1) \quad (7.105)$$

其中 $N(0,1)$ 表示标准正态分布。因此构造出的统计函数为

$$r = \sum_{i=1}^k \frac{(h_{2i} - h_{2i}^*)^2}{h_{2i}^*} \quad (7.106)$$

r 服从 χ^2 分布，自由度为 $k-1$ 。其中 k 是 h_{2i} 与 h_{2i+1} 所组成数字对的数量（ $h_{2i}^*=0$ 的情况不计算在内）， r 越小表示可疑音频含有秘密信息的可能性越大。结合 χ^2 分布的密度函数计算 h_{2i} 和 h_{2i}^* 相等的概率，即音频被隐写的概率

$$P = 1 - \frac{1}{2^{\frac{k-1}{2}\Gamma\left(\frac{k-1}{2}\right)}} \int_0^r \left[\exp\left(-\frac{t}{2}\right) t^{\frac{k-1}{2}-1} \right] dt \quad (7.107)$$

当 P 接近 1，说明可疑音频含有秘密信息。当 P 接近 0，说明可疑音频不含有秘密信息。上述方法只能检测连续 LSB 替换嵌入的秘密信息，如果秘密信息没有嵌满所有的音频样本，而且嵌入位置伪随机地分布于整段音频，比如 Steghide、Stools， χ^2 检测就难以奏效。

LSB 隐写方法还可以通过改变最低几个位平面来进行信息隐藏。比如常用的隐写工具 Hide4PGP 就可以在 16 位的音频样本中嵌入 4 位的秘密信息。 χ^2 分析方法还可以进一步推广到非最低位平面的 LSB 隐写中。

2. RS 分析法

前面介绍过的 RS 分析方法考虑的是图像位平面之间的相关性，而 LSB 信息隐藏会破坏这种相关性。与图像类似，音频信号的各个位平面之间同样具有一定的相关性，因此，RS 分析法同样也适用于音频。

首先将载体 M 分割为相互独立的块，每个块 $G=(x_1, x_2, \dots, x_n)$ ，每个块包含 n 个相邻的样本。可以定义如下函数来描述每一组数据的随机程度

$$f(x_1, x_2, \dots, x_n) = \sum_{i=1}^{n-1} |x_{i+1} - x_i| \quad (7.108)$$

f 值越小，说明音频块的空间相关性越强。定义二轮置换函数，即 $F^2(x) = F(F(x)) = x$ ，另定义两种置换操作： $F_1: 0 \leftrightarrow 1, 2 \leftrightarrow 3, \dots, 254 \leftrightarrow 255$ 为 $2i$ 与 $2i+1$ 的转换关系； $F_{-1}: -1 \leftrightarrow 0, 1 \leftrightarrow 2, \dots, 255 \leftrightarrow 256$ 为 $2i-1$ 与 $2i$ 的转换关系。可写作

$$F_{-1}(x) = F_1(x+1) - 1 \quad (7.109)$$

另外, 定义 $F_0(x)$ 函数表示不变关系 $F_0(x)=x$ 。于是, LSB 隐写过程可表述为: 将待嵌入比特位与样本值的最低位相比较, 如果相同, 应用 F_0 保持 LSB 不变; 如果不同, 应用 F_1 改变样本值。

对载体 M 做 F_1 变换, 定义

$$R_+ = \frac{\sum P | f(F_1(P)) > f(P) }{\sum P}$$

$$S_+ = \frac{\sum P | f(F_1(P)) < f(P) }{\sum P}$$
(7.110)

式中分母表示所有分块的数目, 分子表示满足条件的分块数目。同样, 对载体 M 做 F_{-1} 变换, 定义

$$R_- = \frac{\sum P | f(F_{-1}(P)) > f(P) }{\sum P}$$

$$S_- = \frac{\sum P | f(F_{-1}(P)) < f(P) }{\sum P}$$
(7.111)

如果 M 为原始载体, 无论应用 F_1 变换还是 F_{-1} 变换, 从统计上来讲, 都会同等程度的增加音频样本的混乱度, 因此有

$$R_+ \approx R_- > S_+ \approx S_-$$
(7.112)

如果 M 是经过 LSB 隐写的, 应用 F_1 变换和 F_{-1} 变换之后的结果就会有明显的差别。假设嵌入率为 α (平均每个样本含有 α 比特秘密信息), 那么隐写后大约有比例为 $\alpha/2$ 的样本已经做了一次 F_1 变换。若再对其进行 F_1 操作, 假设嵌入率为 β , 则有比例为 $\alpha\beta/2$ 的样本进行了二次 F_1 变换, 回到原来的值。与原始音频相比, 增加了比例为 $(1-\alpha)\beta$ 的样本应用 F_1 变换。这个值随着 α 的增大而减少, 即 R_+ 和 S_+ 间的差距随着嵌入率上升而下降。但对隐写对象再进行 F_{-1} 操作, 部分样本经过一次 F_1 变换和一次 F_{-1} 变换, 不会变回原来值, 只会更进一步增大混乱程度, 故 R_- 和 S_- 之间的距离不会随着嵌入率的上升而下降, 所以有

$$R_- - S_- > R_+ - S_+$$
(7.113)

这样, 通过对一个待测音频应用 F_1 和 F_{-1} 操作, 根据 R_+ 、 S_+ 、 R_- 和 S_- 这四个值之间的关系就可以判断音频是否采用 LSB 方法隐写。

RS 分析方法利用参数 R_+ 、 S_+ 、 R_- 和 S_- 的不对称性, 还能进一步估计出嵌入率。对某一待测音频进行测试, 假设嵌入率为 α , 根据上述算法计算出一组 R_+ 、 S_+ 、 R_- 和 S_- , 再对待测音频的所有 LSB 用 F_1 翻转, 再计算出一组 R_+ 、 S_+ 、 R_- 和 S_- , 将这 4 对值连线并延长至交点, 如图 7.15 所示, 其中 R_+ 和 S_+ 相交在嵌入率为 1 的地方, 而嵌入率为 0 时 $R_+ \approx R_-$, $S_+ \approx S_-$, 因此可以用下式来估计嵌入率

$$\hat{\alpha} = \frac{L_1}{L_1 + L_2}$$
(7.114)

3. 抽样对分析法 (SPA)

抽样对分析方法^[122]是对音频信号的抽样对的有限状态机进行分析的隐写分析方法。SPA 方法根据不同抽样对的值得到了抽样对的集合, 经过分析 LSB 隐写之后抽样对子集之间的状态转换关系, 建立了一个有限状态机, 并可以从该有限状态机中求得嵌入率 p 。

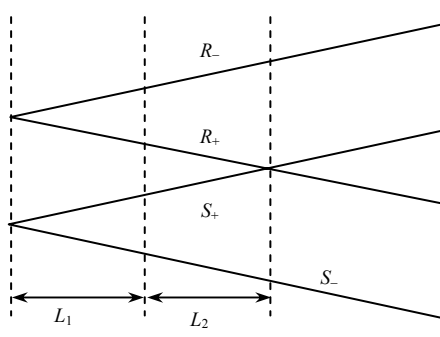
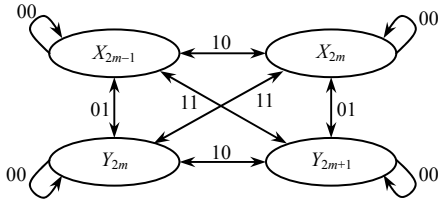


图 7.15 用 RS 方法估计嵌入率示意图

用 s_1, s_2, \dots, s_N 表示音频信号的样本序列, $P = \{(s_i, s_j) \mid 1 \leq i, j \leq N\}$ 表示抽样对集合。 $D_n = \{(u, v) \in P \mid |u - v| = n\}$ 表示 P 的子集, n 为整数。对每个整数 m , 定义: $C_m = \{(u, v) \in P \mid \lfloor u/2 \rfloor - \lfloor v/2 \rfloor = \pm m\}$, 其中 $\lfloor \cdot \rfloor$ 表示向下取整。显然, D_{2m} 包含在 C_m 中, 但是 D_{2m+1} 包含在 C_m 和 C_{m+1} 中。将 D_{2m+1} 分成两个子集 X_{2m+1} 和 Y_{2m+1} , 其中 $X_{2m+1} = D_{2m+1} \cap C_m$, $Y_{2m+1} = D_{2m+1} \cap C_{m+1}$, $0 \leq m \leq 2^{b-1} - 1$ 。假设

$$E\{|X_{2m+1}|\} = E\{|Y_{2m+1}|\} \quad (7.115)$$

该式为抽样对分析法的一个关键的假设。定义 LSB 嵌入过程对抽样对可能产生的置换操作: $\pi = \{00, 01, 10, 11\}$, 0 表示抽样对中对应的样本保持不变, 1 表示抽样对中对应的样本最低位取反。显然, C_m 的统计值在隐写前后保持不变, 而 X_{2m-1} 和 Y_{2m+1} 在 C_m 的子集 X_{2m-1} 、 Y_{2m+1} 、 X_{2m} 和 Y_{2m} 之间转换, 其中 $D_{2m} = X_{2m} \cup Y_{2m}$ 。由此, 可以建立如图 7.16 和图 7.17 所示的有限状态机, 通过统计 LSB 隐写前后子集的变化, 可以估计嵌入率。

图 7.16 C_m 的有限状态机图 7.17 C_0 的有限状态机

设 A 为抽样对集合的任意子集 $A \subseteq P$, 用 $\rho(\pi, A)$ 表示在 LSB 嵌入后 A 中的抽样对被 π 置换的概率, 令 p 为隐写嵌入率, 则经过 LSB 隐写后样本值改变的概率为 $p/2$ 。假设 LSB 嵌入的秘密数据随机分布在载体信号中, 则我们能够得到: ① $\rho(00, P) = (1-p/2)^2$; ② $\rho(01, P) = \rho(10, P) = p/2(1-p/2)$; ③ $\rho(11, P) = (p/2)^2$ 。

通过推导, 最终可以得到下列的二次方程

$$\frac{(|c_m| - |c_{m+1}|)}{4} p^2 - \frac{(|D'_{2m}| - |D'_{2m+2}|) + 2(|Y'_{2m+1}| - |X'_{2m+1}|)}{2} p + |Y'_{2m+1}| - |X'_{2m+1}| = 0 \quad (7.116)$$

其中 $m > 0$ 。类似的, 对于 $m=0$, 有

$$\frac{(2|c_0| - |c_1|)}{4} p^2 - \frac{(2|D'_0| - |D'_2|) + 2(|Y'_1| - |X'_1|)}{2} p + |Y'_1| - |X'_1| = 0 \quad (7.117)$$

由上两式即可解得嵌入率 p 。

7.5.5 通用音频隐写分析方法

虽然目前针对 LSB 的隐写分析方法, 不仅能够判断出待测媒体是否含有秘密信息, 还能估计出秘密信息的长度, 但是在大多数情况下, 我们只能得到隐写体, 而无法得知可疑媒体可能通过哪种隐写方法生成的, 因此通用隐写分析方法有着更重要的意义。

1. 基于音频质量测度的隐写分析方法

Ozer 等^[123]提出了一种基于音频质量测度的通用音频隐写分析方法。他们通过对包括感知域、时域和频域在内的 19 种有关音频质量测度进行了调查, 这些测度原来是用于语音/音频信号的音质评价, 尤其是对压缩编码算法的重建质量评价。由于隐写引起了音频质量的下降, 所以它们可以用来揭示隐蔽信息的存在性。他们通过提取一系列音频质量测度的统计量组成区分是否隐写的统计特征, 然后对特征进行筛选, 去除某些对隐写不敏感的特征, 最后通过分类器进行分类。调查过程如下: 利用小波去噪或其他的去噪方法对待测音频进行估计, 计算待测音频与估计音频之间的质量测度, 发现未隐写的音频载体与嵌入秘密数据的音频的质量测度是不相同的。一般来说, 一个平滑信号与其估计之间的测度小于含噪信号与其去噪信号之间的测度。因此, 采用音频质量测度作为特征向量能够反映出音频是否经过隐写。可供选择的音频质量测度见表 7.3。

Ozer 等人分别采用了方差分析法 (Analysis of Variance) 和 SFS (Sequential Floating Search Method) 法对音频质量测度特征进行了筛选。实验中选取的隐写方法有: 直接扩频法 (DSSS)、跳频扩频法 (FHSS)、基于听觉系统的离散余弦算法 (DCTwHAS)、回声隐藏算法 (ECHO), 还包括隐写软件 Stools 和 Steganos。然后分别采用了多元线性回归和支持向量机 (SVM) 两种分类器进行分类。该隐写分析方法的流程如图 7.18 所示。

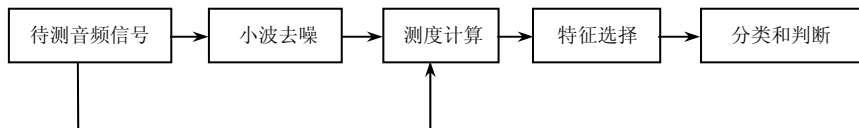


图 7.18 隐写分析流程图

2. 基于直方图特征函数统计矩的隐写分析方法

前面图像通用隐写分析中指出, 将直方图通过傅里叶变换到频域, 即直方图的特征函数 (HCF), 比空域具有更高敏感度。将小波系数直方图进行傅里叶变换, 转换至频域, 再计算频域的高阶统计矩作为特征来对图像进行隐写分析。秘密信息嵌入过程, 相当于载体直方图与秘密信息的概率密度函数进行卷积, 也相当于对载体直方图进行了一次低通滤波, 所以, 在隐写前后直方图频域上的统计量肯定发生了变化, 可以通过捕捉这个变化来发现隐写的行为。这一特性完全可以移植到音频通用隐写分析中, 在此不再赘述。

3. 基于线性预测的隐写分析方法

众所周知, 在一段较短的时间内 (一般为 20ms), 声音信号一般是相对稳定而且前后相关的, 其各频率成分也在短时间内相对稳定且前后相关。各种隐写算法, 在对音频进行隐写的时候, 不仅会影响音频信号的统计特征, 还对其短时间内的前后相关性产生影响。通常, 可以采用小波变换来考察信号各频率成分在时间上的前后相关性。小波分解的不同分辨率分别对应音频信号的不同频带, 分别考虑各个频带中小波系数的相关性, 提取线性预测误差的统计特征能够反映出这种相关性的改变。

线性预测的基本原理是: 音频信号的当前估计值 $\hat{x}(n)$, 可以用前 P 个信号 $x(n-1)$,

$x(n-2), \dots, x(n-P)$ 的加权线性组合来逼近, 即

$$\hat{x}(n) = \sum_{m=1}^P \alpha_m x(n-m) \quad (7.118)$$

由此估计所引入的误差为

$$e(n) = x(n) - \hat{x}(n) = x(n) - \sum_{m=1}^P \alpha_m x(n-m) \quad (7.119)$$

线性预测的基本问题就是估计出一组合适的预测器系数 α_m , 使得均方误差最小化, 即

$$E[e^2(n)] = E\left\{ \left[x(n) - \sum_{m=1}^P \alpha_m x(n-m) \right]^2 \right\} \quad (7.120)$$

为使得 $E[e^2(n)]$ 最小, 对 α_m 求一阶偏导数, 并令其为零, 可以得到一组关于 α_m 的方程, 解此方程组就可得到预测器系数 α_m , 可以用 Levinson-Durbin 递推方法进行快速计算。

这样一来, 一种可行的通用隐写分析方案如下: 首先用 Haar 小波对音频信号进行 4 级小波分解, 然后在每个子带上对小波系数进行线性预测, 提取各子带线性预测误差的均值、方差、偏度和峰度, 再加上各子带小波系数的均值、方差、偏度和峰度作为统计特征。实验表明, 该方案对几种隐写工具如 Steghide、Hide4PGP 和 Stools 具有较好的检测率。

7.6 视频隐写分析技术

现阶段隐写分析研究集中于静止图像。而针对视频信息隐藏的分析技术发展相对缓慢, 这一方面是由于视频信息隐藏及其分析技术需要具备视频编解码系统的研究背景; 另一方面是目前只有很少成熟的视频信息隐藏软件被公开。但数字视频作为未来网络信息资源的重要组成, 基于视频资源的信息隐藏及其隐藏分析技术正逐步成为信息隐藏领域的研究重点。本节首先介绍视频隐写分析的基本框架, 然后介绍视频隐写分析的特点, 接着介绍视频隐写分析的评估指标, 最后介绍两种强针对性视频隐写分析方法。

7.6.1 视频隐写分析基本原理及框架

视频隐写分析的基本思想是: 秘密信息的嵌入虽然不改变视频序列的感观效果, 却会在一定程度上无可避免地造成原始载体视频数据的某些统计特性发生变化, 基于这个特点, 视频隐写分析系统通过分析视频信息量大小与统计特性偏差之间的对应关系来得出可疑性决策, 以至估计出秘密信息在视频序列中的隐藏位置、强度等。

视频隐写分析技术可表述为如下统计假设检验模型

$$F(S_k) = \begin{cases} \text{TRUE} & X_k = C_k + f(C_k, W_k) + \text{noise}, k = 1, 2, \dots, N \\ \text{FALSE} & X_k = C_k + \text{noise}, k = 1, 2, \dots, N \end{cases} \quad (7.121)$$

其中, $F: X \rightarrow \{\text{TRUE}, \text{FALSE}\}$ 为检验函数, X_k 为待检测视频的第 k 帧, C_k 表示载体视频的第 k 帧, W_k 表示嵌入到载体视频第 k 帧中的秘密信息, N 表示该视频序列的长度, $S_k = f(C_k, W_k)$ 表示经某视频隐写算法隐写得到的结果, 该算法与原始视频序列以及嵌入的秘密信息有关, noise 表示噪声。

视频隐写分析的通用框架可以用图 7.19 来表示。首先输入待测视频序列, 对其进行特征提取, 根据提取的特征向量是否被改变以及改变程度的大小来判断视频中隐藏了秘密信息与否。找出对视频隐写敏感或是能反映秘密信息存在的统计特征的特征提取过程

是实现视频隐写分析的关键，一个好的特征向量要能以高准确度和低误差率进行载体视频与隐写视频的正确检测，并且还要不依赖于视频序列的大小，格式以及压缩码率等。接下来进行的判别过程则是根据提取的特征向量，应用各种分类器如支持向量机对待测视频序列进行归类、判断。修正过程则通过观测判别结果的好坏程度来适当修改提取的特征、判别系数或阈值，最终达到正确判断或提高判别的准确性的目的。具备了视频被动隐写分析这个基础和前提之后，可以尝试主动视频隐写分析。主动视频隐写分析一方面可以通过攻击混淆等手段阻止隐蔽通信的进行，另一方面也可以从视频序列中估计出秘密信息所在的具体位置，然后提取出加密过的密文并进一步破译出秘密信息。目前视频隐写分析技术主要针对视频被动隐写分析。

7.6.2 视频隐写分析特点

尽管视频序列可以视为一系列具有高相关性的静止图像在时间域上的有序组合，但由于视频具有一些不同于图像的特殊属性使得视频隐写与图像隐写之间的区别较大，相应地，视频隐写分析与图像隐写分析也必然在某些方面存在一定的差异，因此视频隐写分析技术在借鉴静止图像隐写分析相关技术的同时，还具有其自身独有的特点。

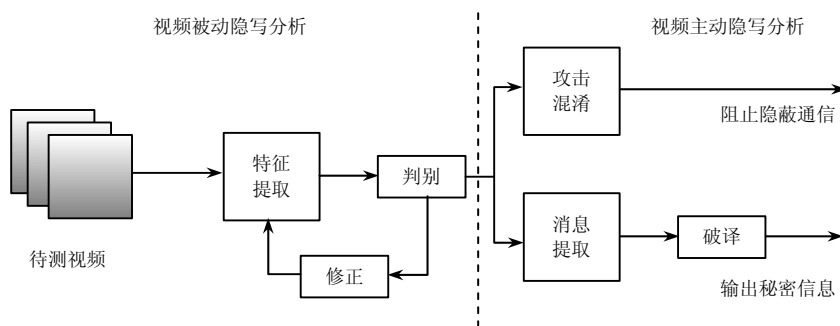


图 7.19 视频隐写分析通用框架

(1) 视频隐写分析需要考虑以数字视频为载体的隐写虽具有绝对大的可用载体空间，但嵌入比率却相对较小。由于数字视频从空间域扩展到了时空域，其具有的数据量必然远远超过一幅静止图像的数据量，同时视频隐写不仅可以利用空域冗余嵌入秘密信息，还可以利用时域冗余进行信息隐藏，从而可知数字视频的可用载体空间是绝对大的。然而，不可忽视的是，数字视频可嵌入比率却相对较小。一方面，为了方便传输与存储，数字视频大多经过了高压缩比的视频压缩编码过程，这使得视频数据量在大幅度降低的同时，也压缩了适宜进行秘密信息嵌入的视频时空域冗余空间；另一方面，为了保证隐写的安全性，在视频载体中选取的嵌入点也不能太多，因为嵌入点越多即意味着可用于隐写分析的载体可用数据总数越大，隐写系统也就越不安全。视频的这个特点使得视频隐写分析要在大数据量的视频流中检测相对小嵌入量的秘密信息，相比图像隐写分析其难度将大大增加。

(2) 视频隐写分析除了可以充分利用视频帧内的空域或变换域相关性进行统计分析之外，还可以利用视频帧之间特有的时域上的高相关性进行隐写分析。尽管视频压缩较大程度上消除了空域冗余，使得隐写导致的空域以及变换域的变化不明显，但对于时域统计特性的变化却很难控制，从而为视频隐写分析提供了一个强有力的分析手段。目前大多数视频隐写分析算法都是基于时域相关性进行隐写分析的。

(3) 视频隐写分析算法需要结合视频编解码系统进行设计。这是由于大多数视频序

列都需要经过压缩编码,从而以视频为载体的隐写算法对视频编解码系统都有很强的依赖性。根据这个特点,可以通过分析视频编解码系统哪些模块可能嵌入了秘密信息,设计出具有强针对性的视频隐写分析方法。若隐写算法独立于视频编解码系统之外,那么压缩过程即相当于对该隐写方式的攻击,为了使得接收方能够正确恢复出秘密信息,则不得不付出大幅度削减实际嵌入数据量的代价。

7.6.3 视频隐写分析算法的评估指标

数字视频被动隐写分析的目的就是要正确判决视频载体是否嵌入了秘密信息。在视频隐写分析中可能出现的判决方式如图 7.20 所示。

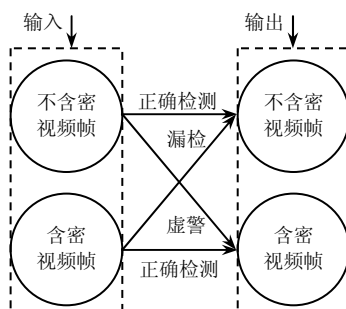


图 7.20 视频隐写分析中可能出现的不同判决方式

将原本不含秘密信息的视频帧判决为隐写视频帧,称之为虚警 (False Positive); 而把原本含有秘密信息的视频帧判决为非隐写视频帧,则称之为漏检 (False Negative)。视频隐写分析算法可从正确检测率,也即检测精度 (Accuracy, ACC), 击中率 (True Positive Rate, TPR), 虚警率 (False Positive Rate, FPR), 漏检率 (False Negative Rate, FNR) 四个方面来衡量分类性能,定义分别如公式 (7.122) ~ (7.125) 所示。视频隐写分析要求在尽量降低虚警率和漏检率的条件下获得最佳检测率。若某视频隐写分析算法能使得虚警率和漏检率都较低,则表明该算法检测性能较好,但虚警率和漏检率又是一对矛盾体,在降低虚警率的同时,漏检率必然会增大,反之亦然。

$$ACC = \frac{TP + TN}{P_1 + N_1} \quad (7.122)$$

$$TPR = \frac{TP}{P_1} \quad (7.123)$$

$$FPR = \frac{FP}{N_1} \quad (7.124)$$

$$FNR = \frac{FN}{P_1} \quad (7.125)$$

式中, P_1 表示隐写视频帧帧数, N_1 表示非隐写视频帧帧数, TP 表示隐写视频帧被正确检测的帧数, TN 表示非隐写视频帧被正确检测的帧数, FP 表示虚警的帧数, FN 表示漏检的帧数。视频隐写分析算法的性能还可以采用 ROC (Receiver Operating Characteristic) 曲线进行衡量。ROC 曲线,又称为相关操作特征曲线,它以 FPR 作为横坐标,以 TPR 作为纵坐标,直观地反映了 FPR 与 TPR 之间的关系。ROC 曲线越向左上角偏,其曲线下的面积 AUC (Area under the Curve) 越大,就说明该分类器的分类性能越好。

7.6.4 强针对性视频隐写分析算法

目前图像隐写分析是数字媒体隐写分析的主要研究领域，而视频隐写分析的研究还尚未完全展开。一方面由于视频帧在时域上的差异导致寻找对隐写敏感的统计特征成为难题，但也正是因为视频特有的时域上的压缩特性，使得在此基础上挖掘视频隐写分析算法具有广阔的前景；另一方面，至今尚未有较成熟的视频隐写软件出现；再就是融合视频压缩编码标准的视频隐写存在很多不确定因素，使得视频隐写分析实际执行起来要比图像隐写分析复杂很多，并且对于该课题的研究者来讲还必须具备视频编解码系统的相关知识，这又让很多人止步于此。现有的视频隐写分析算法主要基于视频时域上的高相关性，充分利用相邻帧所具有的边信息来实施载体视频与载密视频的有效分类。下面根据所针对隐写方式的不同来分类讨论视频隐写分析算法。

1. 针对扩频隐写

Budhia 等^[124]在假设各视频帧 F_i 具有相同的方差和均值，帧间相关系数遵循一阶马尔可夫模型，秘密信息 W_i 服从均值为 0 的高斯分布，且 F_i 、 W_i 之间相互独立的条件下，提出了一种基于**时域帧平均**（Temporal Frames Averaging, TFA）共谋攻击的视频隐写分析算法。TFA 算法以帧 F_i 为中心，应用窗口内相邻 $2L+1$ 帧线性共谋来获得载体或隐写视频帧的估计帧 \hat{F}_i ，其定义如下

$$\hat{F}_i = \begin{cases} \frac{1}{2L+1} \sum_{k=1}^{2L+1} F_k & 1 \leq i \leq L \\ \frac{1}{2L+1} \sum_{k=i-L}^{i+L} F_k & L < i \leq N-L \\ \frac{1}{2L+1} \sum_{k=N-2L}^N F_k & N-L < i \leq N \end{cases} \quad (7.126)$$

图 7.21 给出了基于 TFA 的视频隐写分析框架。 \hat{W}_i 为 TFA 共谋攻击得到的估计秘密信息。该文在秘密信息具有高斯特性的假设下，从估计帧与原始帧作差后的**视频残差帧**（Prediction Error Frame, PEF）中提取出了衡量高斯程度的三个统计特征：峰度（Kurtosis）、熵（Entropy）和四分位数（The 25th Percent）。其中，前两者定义为

$$\text{kurtosis} = \frac{1}{\sigma^2 N_0} \sum_{\forall i} (x_i - \mu)^4 \quad (7.127)$$

$$\text{entropy} = - \sum_{\forall i} p_x(i) \log p_x(i) \quad (7.128)$$

式中 x_i 表示 PEF 中某点像素值， σ 与 μ 分别代表其方差与均值， $P_x(i)$ 为 PEF 中某像素值出现的概率， N_0 为单个视频帧像素点的个数。峰度用来度量一个给定分布的“形状”，若是嵌入了类高斯秘密信息，其峰度将较接近于 3。熵用来反映分布的“随机性”，由于隐写破坏了视频的时空相关性，使得 PEF 中零值减少，从而载密视频的熵要大于载体视频。四分位数定义为残差帧直方图 25% 的点所在区间的值的下限，它反映了给定分布的扩散程度，隐写视频的该值要比载体视频的大。最后利用最近邻（KNN）分类器实现针对在视频每帧每像素均进行嵌入的扩频隐写算法的检测。需要注意的是，简单的 TFA 线性共谋算法仅适用于检测运动缓慢的视频序列。

Pankajakshan 等^[125]强调感兴趣的信号是经运动补偿后的预测误差，因此提出了一种基于**亚像素运动补偿**（Sub-pixel Accuracy Motion Compensation Prediction, SPAMCP）的

视频隐写分析方法。该方法假设通过扩频调制嵌入到视频帧中的秘密信息是运动不一致且独立同分布的高斯信号。考虑到对于存在局部运动或是摄像机引起的全局运动的视频,因运动估计不够精确在 PEF 中残留的载体能量不可忽略,而基于块的运动补偿无法精确捕捉视频的运动,所以该算法为尽量降低残留载体能量对于分类的干扰,采用亚像素运动估计及空域插值得至 IJPEF。其中 P 帧基于前向运动估计, B 帧基于双向运动估计。SPAMCP 方法较好地捕捉了视频的运动,使得基于 SPAMCP 的共谋方式能够较大幅度上去除 PEF 中残留的载体干扰信息,增大载体视频与载密视频问的差异。同时在提取分类用的特征方面,没有简单地用高斯特征来衡量,而是通过观测 P 帧或是 B 帧的 PEF 局部方差直方图是否符合如下 Gamma 分布来进行判定

$$f(x) = \begin{cases} \frac{b^{-a}}{\Gamma(a)} x^{a-1} e^{-\frac{x}{b}} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (7.129)$$

$$\Gamma(a) = \int_0^{\infty} t^{a-1} e^{-t} dt \quad (7.130)$$

式中 a 是形状参数, b 是尺度参数。 a, b 均大于 0。判定方法为:首先去除掉 PEF 局部方差大于 100 的值,再比较 Gamma 分布的形状参数 a 是否大于某一阈值就能有效区分载体视频与隐写视频。该算法性能要优于上述各种方法,并能有效检测包含较迅速或复杂运动的视频,同时还能根据载密视频 PEF 局部方差增大或不变来分辨秘密信息是否是运动一致的,而前面几种视频隐写分析算法是无法抵抗运动一致性隐写的。

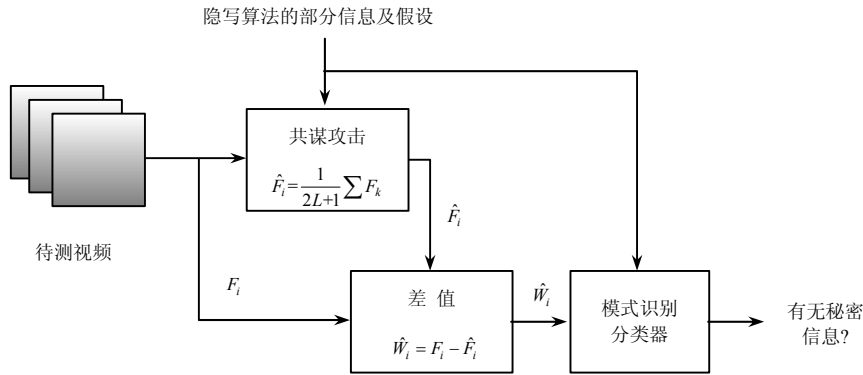


图 7.21 基于 TFA 的视频隐写分析框架

刘滨等^[126]认为不论是何种形式的共谋总会在一定程度上使得 PEF 残留有载体噪声,那么若只是简单地将秘密信息作为相邻帧差别的主要成分来进行特征提取及分类的方法在检测含局部运动的视频时性能必然会显著下降,因此他们将经过 TFA 线性共谋后得到的 PEF 中的秘密信息成分(高斯噪声)及局部运动成分(非高斯噪声)建模为双模噪声,提出一种基于原始视频帧与估计帧分块相关度及非运动块高斯特征提取的视频隐写分析算法。该算法中的双模噪声建模如下。

模型一:视频序列较平稳时,秘密信息为 PEF 中噪声主要成分,双模噪声的概率密度函数可表示为

$$P_1(x) = \frac{1}{\sqrt{8\pi}\sigma} \left[\exp \frac{(x+b)^2}{-2\sigma^2} + \exp \frac{(x-b)^2}{-2\sigma^2} \right] \quad (7.131)$$

噪声绝对值的均值为

$$E[x] = b \left[2\phi\left(\frac{b}{\sigma}\right) - 1 \right] + \sigma \sqrt{\frac{2}{\pi}} \exp\left(\frac{-b^2}{2\sigma^2}\right) \quad (7.132)$$

模型二：视频运动较剧烈时，即 PEF 中存在局部运动成分，双模噪声的概率密度函数可表示为

$$P_1(x) = \frac{1}{\sqrt{8\pi\sigma\phi(b/\sigma)}} \exp\left[\frac{[x - b\operatorname{sgn}(x)]^2}{-2\sigma^2}\right] \quad (7.133)$$

噪声绝对值的均值为

$$E[x] = b + \frac{\sigma}{\sqrt{2\pi\phi(b/\sigma)}} \exp\left(\frac{-b^2}{2\sigma^2}\right) \quad (7.134)$$

由于去除了存在局部运动的非高斯噪声分块，仅对嵌入操作敏感的非运动块提取高斯统计特征，因此该算法在一定程度上降低了局部运动的影响，改善了分类效果，并且能较快速实现载密视频帧的检测。

2. 针对视频隐写软件—MSU StegoVideo

视频隐写软件 MSU StegoVideo 是一款结合视频内容进行隐写的软件，它采用高鲁棒性的扩频调制隐写方式，通过冗余度参数与嵌入强度参数分别来控制扩频系数以及叠加调制强度，能有效抵抗各种攻击以及二次压缩编码。

苏育挺等^[127]通过反向分析得出了 MSU 采用的是类似棋盘格分布的嵌入模式，该模式以 32×32 像素块为基本单位，对其中的 4 个 16×16 块以相互交错的调制幅度进行秘密信息的嵌入。这种嵌入模式在增强了 MSU 鲁棒性的同时，也引入了一种新的整体分布不均匀的块效应，而视频压缩编码量化导致的块效应整体分布却是趋于均匀的，正是基于这两种块效应分布之间的差异，苏育挺等人设计了一种根据块边界差值来检测 MSU 隐写的视频隐写分析算法。该算法能够有效检测不同嵌入强度及不同压缩码率下的 MSU 隐写，但由于判决门限易受序列长度、载体视频内容、编码质量控制策略等的影响。为了提高检测性能，如何选择合适的判决门限还有待进一步研究。2008 年 Zhang 等^[128]又提出了一种基于帧间模式检测 MSU 隐写的算法，该算法通过计算存在于相邻帧差值图像中的特殊模式数量是否达到一定阈值来判断视频中是否隐藏了秘密信息，文中采用的特殊模式是通过计算差值图像中每 32×32 像素块中 4 个 16×16 像素块的 DC 系数的符号是否存在一致性来实现的。该算法检测效果稍好于前一种，进一步的改进方向可从加强 MSU 隐写视频的帧内与帧间信息统计特征的研究着手。

Liu 等^[129]针对 MSU 隐写提出了一种基于扩展马尔可夫特征与变换域联合分布特征的视频隐写分析算法。该算法对每个测试视频提取出了 1944 个特征，包含 DCT 块内、DCT 块间以及 DWT 子带各四个方向的马尔可夫特征，以及它们的联合分布特征，为提高运算效率，采用方差分析法 (ANOVA) 对选取的特征进行降维处理，再用 SVM 实现有效分类。

Wu 等^[130]在苏育挺等人对 MSU 研究的基础上，认为 MSU 隐写主要影响视频帧的低平面比特位，并且发现用于建立运动矢量以及图像边缘的区域的 MSU 嵌入模式并不严格遵循 16×16 的子块分割方式，这可能会造成块边界分布的改变，从而使得基于块边界差值的检测算法出现误判，因此该文提出了一种基于不连续系数的视频隐写分析方法。不连续系数 P_0 定义如下

$$P_0 = \frac{D_{16}}{D_{32}} \quad (7.135)$$

其中, D_{16} 与 D_{32} 分别为检测单元分别为 16 像素与 32 像素游程的块度量因子, 定义如下

$$D_{16} = \frac{1}{n_{16}} \sum_0^{16} \frac{|\text{sum}_0 - \text{sum}_1|}{16} \quad (7.136)$$

$$D_{32} = \frac{1}{n_{32}} \sum_0^{32} \frac{|\text{sum}_0 - \text{sum}_1|}{32} \quad (7.137)$$

其中, sum_0 和 sum_1 表示沿着帧数据低平面比特位的水平与垂直方向扫描 16 像素或 32 像素得到的 0 或 1 的数目, n_{16} 和 n_{32} 则分别表示采用 16 像素或 32 像素游程的检测单元的数目。

对于隐写视频来说, $D_{16} \approx 1$ 、 $D_{32} \approx 0$, 从而使得 $P_0 \gg 1$, 而非隐写视频的编码单元是 8×8 的像素块, $D_{16} \approx D_{32}$, 从而使得 $P_0 \approx 1$ 。通过对不连续系数设置合适的阈值, 可以实现较其他针对 MSU 的视频隐写分析算法更优的检测性能。

7.6.5 视频盲隐写分析算法

Pankajaksha 提出了一种时空域双重预测的视频隐写分析算法^[131]。为了降低时域及空域残留非秘密信息载体噪声对分类的影响, 该算法首先采用运动估计补偿去除视频序列的时域冗余, 并利用如下空域预测方法对时域预测残差帧去除空域相关性

$$\hat{x} = \begin{cases} \max(a, b) & c \leq \min(a, b) \\ \min(a, b) & c \geq \max(a, b) \\ a + b - c & \text{其他} \end{cases} \quad (7.138)$$

其中, \hat{x} 为 x 的预测值。像素 a , b , c , x 的相对位置如图 7.22 所示。再对经时空域预测得到的 PEF 进行 3 级离散小波变换, 提取小波域每个子带特征函数的前三阶矩, 最后用得到的 39 维特征向量去训练模式识别分类器实现秘密信息的正确检测。该算法改善了低嵌入率下视频隐写分析的性能, 但对运动较快或是含非平移运动的视频检测效果较差。

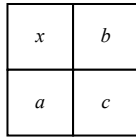


图 7.22 像素 a , b , c , x 的相对位置

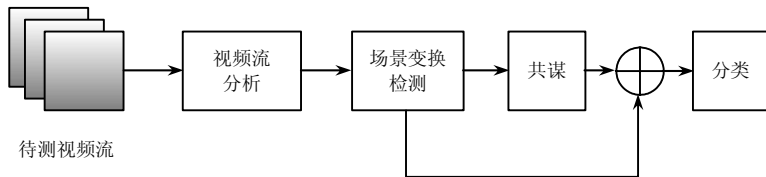


图 7.23 视频盲隐写分析框架

针对压缩域视频隐写, 刘镔等^[132]提出了一种基于帧间相关性的视频盲隐写分析 (Inter-Frame based Correlation Steganalysis, IFCS), 其算法框架如图 7.23 所示。考虑到不同场景下的视频帧特征差异较大, 因此在共谋之前, 文中利用压缩视频的 DC 系数进行了场景切换点的检测, 检测函数如下式所示

$$\text{HD}(I_i, I_{i+1}) = \sum_{k=1}^N \frac{[H_i(k) - H_{i+1}(k)]^2}{[H_i(k) + H_{i+1}(k)]^2} \quad (7.139)$$

I_i 和 I_{i+1} 分别表示视频序列第 i 帧以及 $i+1$ 帧, H_i 和 H_{i+1} 分别表示第 i 以及 $i+1$ 帧的 DC 系数直方图。若相邻两帧 DC 系数直方图的相似度函数值为峰值, 则可确定此处即为所要寻找的场景切换点。刘镔等人采用基于 TFA 的简单线性共谋方法, 窗口大小的选取则根据帧间相关系数来进行自适应调节, 以尽量降低因帧间相关度低引入的共谋噪声。最后提取压缩视频流 DCT 域帧间相关性统计特征、DC 系数、AC 系数作为使用帧间共谋策略的分类用特征向量, 并与场景切换点的信息一并输入到三层前馈非线性神经网络实施载体视频与载密视频的有效分类。在嵌入率大于 30% 时, IFCS 方法检测效果较好。

7.7 本章小结

隐写分析是对隐写术的攻击, 目的在于揭示隐写对象中是否存在秘密信息以致破坏保密通信。本章首先介绍了隐写分析的基本概念、分类和评价指标, 然后分别介绍针对图像、音频和视频载体的隐写分析技术, 其中以图像载体为主。针对每种载体, 主要分专用分析法和通用分析法分别进行介绍。对于图像载体的专用隐写分析, 主要介绍针对空域 LSB 替换的专用隐写分析算法、针对 LSB 匹配隐写的专用隐写分析算法和针对 JPEG 图像隐写的专用分析算法。对于图像载体的通用隐写分析, 主要介绍基本思想、支持向量机分类技术、算法概述和一些典型通用隐写分析算法。对于音频载体, 先介绍专用音频隐写分析的机理和通用音频隐写分析的机理, 然后介绍隐写系统模型和一些典型专用和通用隐写分析方法。对于视频载体, 先介绍基本原理、框架、特点和评估指标, 然后介绍几种典型专用和通用隐写分析方法。

设计可以抵御通用型隐写分析术的隐写算法是大势所趋。通用型隐写分析术主要抓住了载体图像或隐写图像的共性, 将高维的图像表示空间通过特征提取的方法映射到较低维数的特征空间, 当载体图像和隐写图像能被表征在特征空间截然不同的区域时, 隐写算法就被攻破。因此, 如何保持图像的统计特征, 并且在较高的数据嵌入率下实现隐写安全性是隐写术必须解决的问题。而突破通用型隐写分析术的攻击, 必然是重点研究方向。

随着信息隐藏技术的发展, 使用未压缩的音频作为载体的隐写方法逐渐暴露出缺点, 因此, 在研究以未压缩格式音频为载体的隐写与分析的基础上进一步研究以 AAC 压缩音频为载体的隐写与分析方法具有更为实用的意义。目前虽然已经有许多关于基于图像的隐写分析算法和系统, 但在对基于视频的隐写分析算法和系统方面还是很缺乏。至今仍然还没有形成基于视频的隐写分析技术的完善理论或系统。因此, 基于视频的隐写分析技术还有待进一步发展和完善。



习题

1. 请阐述隐写分析技术的基本概念、分类和评价指标。
2. 请用 Matlab 或 C 语言编写一段程序, 基于 Lena 图像来验证 χ^2 检测方法和 R-S 检测方法对 LSB 替换隐写的检测能力。
3. 通过对大量自然图像相邻像素的统计分析, 发现自然图像相邻像素对的像素值之间具有如下关系: (1) 相邻像素对中, 由一奇一偶像素构成的对数与由两奇或两偶像素构成的对数大致相等, 即大致占全部像素对的 50% 左右。(2) 在一奇一偶像素构成的相邻像素对中, 奇像素大于偶像素的对数与偶像素大于奇像素的对数大致相等, 即各占全

部像素对的 25% 左右。即对于原始自然图像, 奇像素大于偶像素的相邻像素对数与偶像素大于奇像素的相邻像素对数的比值等于 1。经过 LSB 替换隐写后生成的隐写图像, 发现奇像素大于偶像素的相邻像素对数与偶像素大于奇像素的相邻像素对数的比值大于 1。请用 Matlab 或 C 语言编写一段程序, 基于 Lena 图像来验证这个现象, 并设计出一种简单的针对 LSB 替换隐写术的隐写分析方法。(可以参考 2007 年秦姣华等人在《系统仿真学报》第 19 卷第 24 期发表的文章《基于相邻像素统计特性的 LSB 隐写分析技术》)

4. 对于一幅灰度图像, 令 $c(i, j)$ 表示图像在位置 (i, j) 的灰度值 ($0 \leq c(i, j) < 256$)。图像像素点 (i, j) 的邻域度定义为 $d(i, j) = |\{(i+u, j+v) | c(i+u, j+v) = c(i, j)\}|$, 其中 $-k \leq u \leq k, -k \leq v \leq k$ 且 u, v 不同时为 0, k 为邻域大小参数, $k=1$ 表示 3×3 的图像邻域, $k=2$ 则表示 5×5 的邻域。也就是说, 图像像素点 (i, j) 的邻域度 $d(i, j)$ 表示邻域内与中心点像素值相同的像素点个数。图像像素点的邻域度体现了该像素跟周围像素相关性, 邻域度越大, 则该邻域内像素的相关性也越强。试用 Matlab 或 C 语言编写一段程序, 基于 Lena 图像 ($k=2$) 来验证经过 LSB 匹配隐写嵌入以后, 邻域度较大的像素数量会减少。试基于邻域度直方图 $h(x) = |\{(i, j) | d(i, j) = x\}|$ 的统计特性, 提出一种针对 LSB 匹配隐写术的隐写分析方法。(可以参考 2011 年的夏槟撰写的湖南大学硕士学位论文《基于图像邻域度的隐写分析方法研究》)

5. JPEG 图像压缩编码过程中, 首先是将图像分为 8×8 大小的不重叠的像素块, 然后对每个像素块进行 DCT 变换, 最后对 DCT 系数进行量化和熵编码。隐写分析在原则上应该选取对秘密信息敏感而对载体图像不敏感的特征。通过研究发现, 将秘密信息嵌入 JPEG 图像后会改变其 DCT 系数的统计分布特性。在图像量化后的 DCT 系数中嵌入秘密信息时, 就会造成 DCT 系数的统计特征发生变化, 我们可以基于这点进行隐写分析。试用 Matlab 或 C 语言编写一段程序, 以 Lena 图像的 JPEG 压缩版本为载体图像, 测试经 F5 隐写算法后的 Lena 图像的 DCT 系数统计特征的变化规律。

6. 为了理解支持向量机, 请求如下分两类问题的分类面方程: 输入 $(0, 1)$ 和 $(0, 0)$ 期望输出 -1, 而对于输入 $(2, 0)$ 和 $(0, 2)$ 则期望输出 1。

7. 设计一种基于空域灰度共生矩阵的空域通用隐写分析方案。灰度共生矩阵是从图像位置 (x_1, y_1) 灰度为 i 的像素 $c(x_1, y_1)$ 出发, 统计与其距离为 δ 、灰度为 j 的像素 $c(x_2, y_2)$ 同时出现的概率而得到的矩阵 $M_{\delta, \theta}(i, j) = \{\Pr(c(x_1, y_1) = i, c(x_2, y_2) = j), x_2 = x_1 + \delta \cos \theta, y_2 = y_1 + \delta \sin \theta\}$, 对于 256 灰度的图像, i, j 的取值范围从 0 到 255。灰度共生矩阵的第 i 行 j 列元素表示图像上所有在 θ 方向, 相隔为 δ , 一个灰度为 i 值, 另一个灰度为 j 值的像素对出现的概率。灰度共生矩阵反映了图像关于方向、相邻间隔和变化幅度的综合信息, 可作为分析图像特征的二阶统计量, θ 的取值一般为 0° 、 45° 、 90° 和 135° 。基于灰度共生矩阵, 可以定义其能量为矩阵所有元素之和; 定义惯性矩为: 位置 i 和 j 乘积的平方乘以相应灰度矩阵元素的平方, 然后对所有位置求和; 可以定义熵为: 每个元素的对数乘以该元素本身, 然后对所有位置求和。除此之外, 还有好多特征可以定义, 请读者自行查阅参考文献。试用 Lena 图像来测试其分别经过 LSB 替换隐写、LSB 匹配隐写后灰度共生矩阵的能量、惯性矩和熵等一系列特征的变化。利用 SVM 技术, 实现基于空域灰度共生矩阵的通用隐写分析方案。

8. 请阐述 LSB 隐写、回声隐藏对音频的影响, 分别概述针对它们的隐写分析特征。

9. 请阐述隐写对哪些音频质量测度会产生影响, 请选择几种测度构成一个特征向量, 利用 SVM 技术实现基于音频质量测度的通用隐写分析。

10. 请阐述视频隐写分析基本原理及框架, 分别叙述一种典型的专用隐写分析方法和一种典型的通用隐写分析方法。

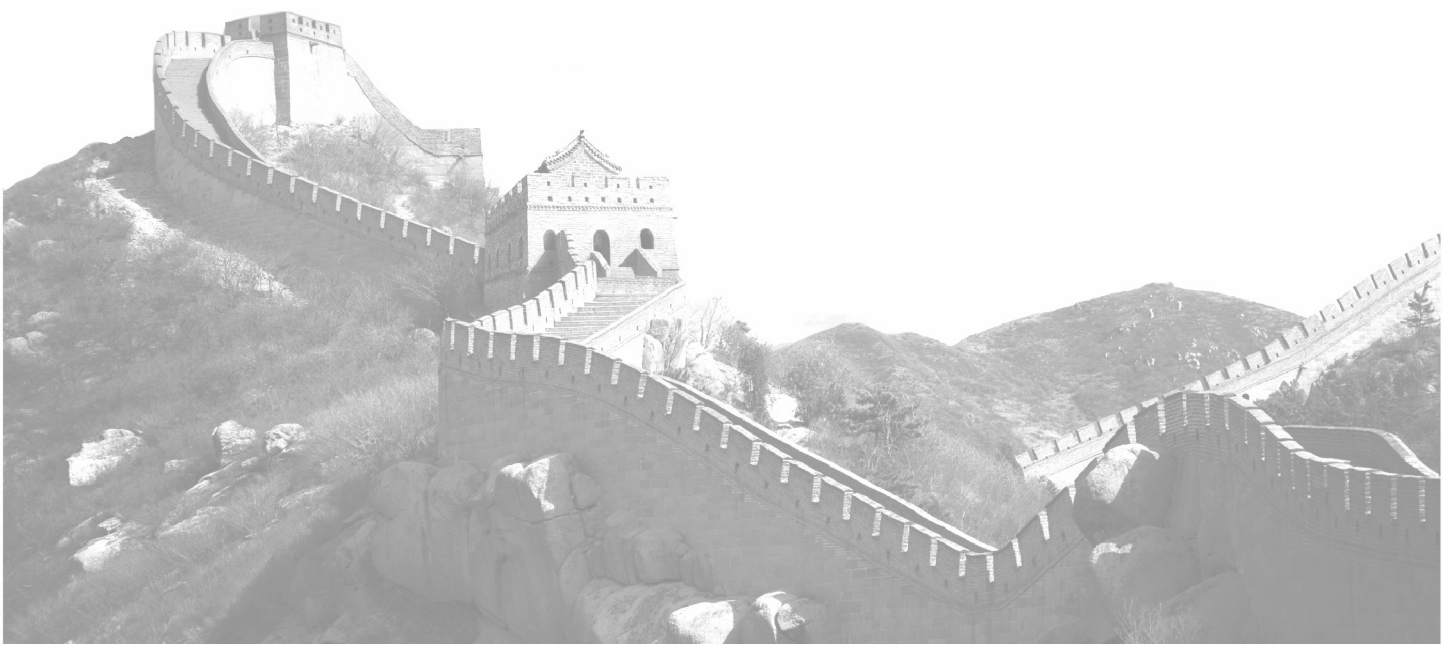
数字水印攻击技术

本章引言

众所周知，若要把数字水印技术真正地应用到实际的版权保护、内容认证等领域，必须考虑的一个重要问题是系统可能受到的各种攻击。当然，不同的应用场合有不同的抗攻击能力要求。抗攻击能力是数字水印系统评测最重要的性能指标，因此有必要系统地了解攻击的种类和常见的算法。本章首先给出攻击的定义、分类和相关概念，然后介绍针对安全性的三种攻击技术，最后介绍针对应用系统而与水印算法安全性及鲁棒性无关的系统攻击技术。

本章重点

- 数字水印攻击技术的相关概念和分类；
- 非授权去除攻击；
- 非授权嵌入攻击；
- 系统攻击。



8.1 数字水印系统的鲁棒性和安全性

8.1.1 鲁棒性

鲁棒性是指含水印产品在经过常规信号处理操作后仍能检测出水印的能力^[5]。对图像的常规处理操作包括空间滤波、有损压缩、打印与复制、几何变形（旋转、平移、缩放及其他）等。视频水印也应同样具有对这些变换的鲁棒性，另外还需具有对磁带转录操作的鲁棒性，以及对受到其他因素影响而引起的帧率变化的鲁棒性。音频水印则需要具有对时域滤波、音频磁带转录以及不等速度回放等操作的鲁棒性。

鲁棒性是版权保护数字水印方案的基本要求，满足下列情况的水印技术通常认为是鲁棒的，即如果为成功删除水印而造成对“图像质量”的损害是过度的，或者攻击者为达到目的而付出过多的时间。但是，鲁棒性定义建立在一个模糊和不确定术语的基础之上。比如如何定义质量？达到什么程度是质量的过度损失？以媒体创造者们的标准来定义质量和鲁棒性是很容易的。这些人将是真正使用水印软件的，所以他们在质量的定义中已经对加入水印造成的破坏做了限制。一个水印嵌入程序不能只靠增加嵌入强度来提高鲁棒性。如果这么做，那它只能用于嵌入引起过度噪声能被接受的领域。在这些领域，水印需要抵抗实质的损害。

并非所有的水印应用都必须具有对所有可能信号处理操作的鲁棒性。一般说来，在水印嵌入操作与检测操作之间可能会存在一些信号处理操作，而水印只需具有对这些信号处理操作的鲁棒性即可。例如，在电视和广播监控中，水印只需在传输过程中存留下来即可。而在视频广播应用中，还须对有损压缩、数模转换、模拟传输（引起低通效应）、添加噪声以及一些水平与垂直的平移具有鲁棒性，用于这种应用的水印通常无需考虑旋转、缩放、高通滤波等经常出现在水印嵌入操作之前以及检测操作之后的处理（易引起质量退化）。例如，用于广播监控的水印无需具有对 VHS 录制的鲁棒性。

在某些情况下，鲁棒性毫无用处甚至极力避免。事实上，水印研究的一个重要分支就是脆弱水印。脆弱水印具有和鲁棒性相反的脆弱特性。例如，用于真伪鉴别的水印就应该是脆弱的，对图像做任何信号处理操作都会将水印破坏掉。在另一类极端应用中，水印必须对任何不至于破坏含水印作品的失真都具有鲁棒性，因为有时无法预料嵌入和检测操作之间的信号处理操作。

8.1.2 安全性

1. 安全性含义

安全性定义为水印能够抵抗有意篡改和恶意攻击的能力^[5]，有别于鲁棒性所涉及的普通信号失真。恶意攻击是指任何有意破坏水印功能的行为。正如计算机安全系统一样，一个系统的安全性与它最薄弱一环的安全性等同。如果攻击者破坏了水印生命周期的任何一个阶段，该攻击就是成功的。因此，作品的拥有者和水印软件必须确保任何一个阶段都非常安全。对水印进行攻击可采用不同的方式，每种应用都需有自己的安全类型和安全级别。

不同的应用对水印系统的安全要求相差甚远。在一些应用当中，由于没人会去篡改或者破坏水印的功用，几乎可以不考虑安全性。比如，在设备控制应用中，水印作为其嵌入内容的附加价值而存在。在这类场合，对水印进行篡改得不到任何好处，虽然水印需要具有对正常操作的鲁棒性，但水印并不需要具备抵抗任何恶意攻击的安全性。

2. 密码和水印密钥

在密码学中, 安全性通常由密钥提供。在对称密钥体制下, 使用给定密钥进行加密的消息只能使用相同的密钥进行解密。利用加密函数 $E_K(\cdot)$ 和密钥 K , 输入明文 m , 可产生密文 m_c , 即

$$m_c = E_K(m) \quad (8.1)$$

密文可通过解密函数 $D_K(\cdot)$ 和相应的密钥 K 解密为明文消息, 即

$$m = D_K(m_c) \quad (8.2)$$

实际上, 水印嵌入和检测也可看作加密与解密操作。水印嵌入通常用嵌入函数 $E(\cdot, \cdot)$ 表示, 输入为消息 m 和原始作品 x , 输出为含水印作品 x^w 。水印检测用检测函数 $D(\cdot)$ 表示, 输入为可疑含水印作品 x' , 输出为消息 m' 。在许多水印系统中, 含水印作品到消息的映射可通过水印密钥 K 来控制。在明检测情况下, 可认为检测密钥中包含原始作品, 从而明检测系统可把每一件作品同一个独立密钥相关联。这样一来, 就可用下面两式刻画所有的水印系统

$$x^w = E_K(x, m) \quad (8.3)$$

$$m' = D_K(x^w) \quad (8.4)$$

这两式与式 (8.1) 和式 (8.2) 非常类似。

在引入密钥的概念之前, 绝大多数加密算法的安全性依赖于对这些算法进行保密。这导致如下两个问题: ① 如果算法安全性遭到破坏, 则必须开发一套全新的加密算法。② 算法不可能公开给别的研究者。这也意味着在算法实际使用之前会存在很多一直未被发现的弱点。如今这些问题基本上可用密钥来解决。在现代加密算法中, 安全性只取决于密钥的安全性, 而不是整个算法的安全性。这样, 加密算法和解密算法便可公开并使用一个密钥进行验证。而在实际使用时可采用另一个不同密钥。而且万一这个密钥被泄露, 也不必更换整个算法, 只需选择一个新的密钥即可。因此, 一个设计完好的密码系统应符合如下标准: ① 公开加密和解密算法不会影响系统安全性; ② 安全性基于密钥的使用; ③ 密钥在一个大密钥空间内选择, 且在此空间中进行穷尽搜索不可行。

人们希望水印算法也能具有同样的标准。但是, 水印算法与密码系统的安全性要求实际上并不相同, 并不能简单地将密码学方法应用于水印技术中。密码只用于防止对消息的非法读写, 因而可用水印中以防止某些形式的被动攻击和伪造行为, 但不涉及水印去除的问题。这个问题类似于军用通信中的信号阻塞。为避免阻塞和干扰, 军方通常利用扩频通信技术使一个窄带信号分布在更宽的频谱范围内。而具体的分布方式对传输器和接收器而言都是保密的。在分布方程未知的情况下, 对手几乎不可能检测和干扰传输。这种分布方程同密码学中的密钥类似, 也是基于加密技术的。水印算法的设计也可使用类似扩频方式的密钥。例如, 可将伪随机噪声模型 (PN 模型) 应用到图像水印算法中。在系统运行时, 嵌入器和检测器必须用相同的 PN 模型。这样, 该 PN 模型, 或者说产生此 PN 模型的种子, 可看作密钥。理想情况下, 若密钥未知, 即使水印算法已知, 也不可能检测出作品中是否存在水印。甚至在密钥只被值得信任的少数人所知时 (即防止对手得知密钥), 对手也不可能在完好保持含水印作品感官质量的前提下成功去除水印。由于在嵌入和检测过程中使用的密钥同密码学中的密钥所提供的安全性不同, 人们经常在水印系统中同时使用这两种密钥。即消息使用一个密钥进行编码以生成水印, 而嵌入则使用另一个密钥。为区分这两种密钥, 分别用生成密钥和嵌入密钥来表示。

3. 公有水印处理和私有水印处理

有时, 读者在文献中会看到公有水印处理 (Public Watermarking) 和私有水印处理

(Private Watermarking) 字样。为了理解和区分这两种水印系统, 引入两组人群: 一组为受信任人群 (Trusted Individuals), 通常指水印设计的受益人; 另一组为公众 (Public), 通常指潜在的对手 (Adversaries)。私有水印处理是指公众在任何情况下都不允许对水印系统进行操作。公有水印处理是指公众只允许进行水印检测操作, 而不允许进行其他操作。在这个意义上, “公有” 和 “私有” 水印处理是指应用对安全性的要求。类似称谓还有**私有水印** (Private Watermarks) 或**私有水印算法** (Private Watermarking Algorithms) 以及**公有水印** (Public Watermarks) 或**公有水印算法** (Public Watermarking Algorithms)。然而, 这种 “公有” 和 “私有” 的称谓显得模棱两可, 因为绝大多数技术都同时适用于这两种类型的应用中。例如: ① 假设公众无法获得原始作品, 那么使用明检测器的系统 (检测操作需要原始作品) 常被称作 “私有” 系统。然而, 确实存在公众能够获得原始作品的公有水印处理系统。假想一个 NASA 网站上有一些空间探测图像, 每一幅原始图像都附有一系列经过各种图像处理技术增强后的版本。每一幅增强图像都嵌有一个标识所经历处理操作的文本水印。同时给定原始图像和某一幅增强图像, 明检测器可以检测出描述增强处理操作的文本串, 在该场景下明检测器用在公有水印处理应用中。② 应用无密钥盲检测的系统, 常称为 “公有” 系统, 它假定公众都知道水印检测的方法。然而, 若算法保密, 则此系统仍可用在私有水印处理中。当然, 这种系统的安全性在实施前很难确定。③ 在应用盲检测的系统中使用密钥, 这对所讨论的两类应用均适合。此时, “公有” 和 “私有” 系统的差别仅在于密钥发放形式的不同。

用 “公有” 和 “私有” 的称谓来描述水印处理技术, 除了存在上述问题之外, 该称谓还极易与 “**公钥密码术** (Public-key Cryptography)” 概念混淆。因此, 下面避免使用 “公有水印” 和 “私有水印” 的称谓。

8.2 数字水印攻击技术的相关概念和分类

8.2.1 攻击方法的分类

本章对攻击的理解是广义的, 即包括针对鲁棒性的常规信号处理操作, 针对安全性的恶意攻击以及针对水印应用系统和法律问题的攻击。据此, 攻击技术可分为三大类:

① 鲁棒性攻击。这种攻击是指含水印作品在检测器之前必须经历的或者可能经历的常规信号处理操作。② 安全性攻击。这种攻击是指攻击者为了政治、经济或军事利益或者纯粹为了恶作剧而对水印算法、水印密钥或者含水印作品所进行的各种恶意攻击。③ 系统攻击。这种攻击不是针对水印算法的鲁棒性和安全性, 而是针对水印应用系统中所涉及到的其他问题 (如硬件设备的安全问题、标准化问题和法律问题)。而狭义上的攻击, 则是指针对安全性的恶意攻击和系统攻击。也有人把鲁棒性攻击归入安全性攻击中的一种。也可以把攻击分为无意攻击 (鲁棒性攻击) 和恶意攻击 (安全性攻击和系统攻击) 两大类。也有人把攻击分成主动攻击和被动攻击两大类。本章着重介绍针对安全性的攻击方法和针对水印应用的系统攻击方法。而对于鲁棒性攻击, 本章隐含地将它归入安全性攻击中 (归入其中的非授权去除攻击)。

安全性攻击^[5]可归纳为三大类: **非授权去除** (Unauthorized Removal)、**非授权嵌入** (Unauthorized Embedding) 和 **非授权检测** (Unauthorized Detection)。非授权去除和非授权嵌入会改动含水印作品, 因而可看作是**主动攻击** (Active Attacks); 而非授权检测不会改动含水印作品, 可看作**被动攻击** (Passive Attacks)。这些攻击的重要性取决于具体应

用。事实上在一些情况下,没有人会恶意攻击水印,水印也无需对任何攻击保持安全性(往往是在水印用作消费者增强产品使用功能的场合)。但对于那些确实需要一定安全程度的场合,弄清各种攻击之间的差别非常重要。

非授权去除是指通过攻击使作品中的水印无法被检测到^[5]。这种攻击通常具有两种形式:去除攻击和掩盖攻击。直觉上讲,水印的去除指被攻击的作品不再会被认为含有水印。一旦水印被去除,即使使用更为先进的检测器也无法检测出水印的存在。注意去除水印并不意味着必须重新构造出不含水印的原件,对手的目的在于制作出在感官上同作品原件几乎一样的复制,却不会被检测器检测出水印。当然原始作品本身也是满足这个条件的作品之一。大多数鲁棒性攻击(压缩、滤波等正常操作)都可归为去除攻击。而在水印的掩盖攻击中,被攻击的作品实际上仍含有水印,但使用现有的检测器很难检测到。通过使用更先进的检测器也许可检测到它的存在。例如,许多图像水印检测器很难检测到经过略微旋转的水印。对手于是可对图像作几乎无法察觉的微小旋转操作,使得经过改动的图像仍具有很高的逼真度。由于检测器对旋转十分敏感,因而无法检测出水印的存在。在这种情况下,仍能利用更为先进的具有旋转校正功能的检测器将水印检测出来。水印的非授权去除还有一种有趣的形式,称作共谋攻击。这种情况下,攻击者设法得到给定作品的数份复制,其中每一份都具有不同的水印,再综合所有版本的复制,产生出一份不含水印的复制。这是交易跟踪应用非常关心的问题,因为交易跟踪需要在每份复制中嵌入不同的水印。实际上,鲁棒性攻击中的几何变形可以归入到掩盖攻击中。这里需要指出,鲁棒性攻击中的任何一个操作都可能变成恶意攻击中的一个重要手段,有时比较难以判定恶意攻击和无意攻击,比如几何攻击,既可以看作鲁棒性攻击,但也是恶意攻击的重要手段。正是这个原因,本章把鲁棒性攻击归入到非授权去除攻击中。

非授权嵌入,亦指伪造(Forgery),即在作品中嵌入本不该含有的非法水印信息^[5]。例如,在真伪鉴别的应用中,人们并不关心对手是否能够构造一个检测不出来的水印,那样做只能恰恰证明作品是假的。相反,如果对手能够进行非授权嵌入,便能够使检测器将伪造的作品错误的鉴别为真。

非授权检测,或者称被动攻击,可以按严重程度分为三个级别^[5]:最严重攻击指对手检测并破译了嵌入的消息。这是非授权解读最直接而且全面的方式。次严重攻击指对手检测出水印,并辨认出每一点印记,但却不能破译这些印记的含义。由于水印同所嵌入的作品相关,对手可能会将其同含水印作品进行对比,从而猜测水印含义。末严重攻击指对手可以确定水印的存在,但却不能够对消息进行破译,也无法分辨出嵌入点。

与水印技术相比,隐写术更关心被动攻击。但在一些应用中,被动攻击问题也显得很重要。例如,设一家广播监控公司免费嵌入水印,并向客户收取监控报告的费用。若对手可解读水印,便可在不投入任何嵌入费用情况下,提供可与监控公司相竞争的业务。

除了上面的攻击分类方法之外,还可以把水印的攻击方法分为四大类:**“鲁棒性”攻击(去除攻击, Robustness Attack or Removal Attack)**、**表达攻击(Presentation Attack)**、**解释攻击(协议攻击, Interpretation Attack or Protocol Attack)**、合法攻击。注意这里的“鲁棒性”攻击和本章理解的鲁棒性攻击含义不太一样。这里的“鲁棒性”攻击是指直接攻击,对应于非授权去除攻击中的去除攻击,目的在于去除水印而不影响载体作品的使用,如一些常用的无恶意信号处理方法:压缩、滤波、缩放、打印和扫描等。表达攻击并不需要去除数字产品中的水印,它通过一定的操作使水印检测器无法检测到水印的存在,这种攻击对应于非授权去除攻击中的掩盖攻击。对于图像载体,它一般通过图像的

集合操作完成，如缩放、平移、旋转、修剪、裁剪、像素交换、重采样、像素插入以及其他一些几何变换等。解释攻击（协议攻击）试图对水印的所有权产生争议，比如一个攻击者试图在一个含水印图像中再次嵌入另一个水印，从而导致所有权争议，这种攻击对应于非授权嵌入攻击。合法攻击对应于系统攻击，可能包括现有的及将来的有关版权和有关数字信息所有权的法案，因为在不同的司法权限中，这些法律可能有不同的解释。这种分类方法没有考虑非授权检测攻击。根据检测情况，也可以将攻击方法归成三类：① 水印无法检测到；② 假水印被检测到；③ 非授权检测。实际上，前两类分别对应非授权去除攻击和非授权嵌入攻击。此外，有些学者把攻击分为如下四类：① 去除攻击；② 几何攻击；③ 密码攻击（Cryptographic Attack）；④ 协议攻击。其中，几何攻击对应非授权去除中的掩盖攻击。密码攻击旨在破坏水印算法中所采用的安全机制以便找到一种去除水印或嵌入假水印的目的。例如**强力攻击**（Brute-Force Attack）的目的就是对嵌入的秘密信息进行强力搜索，而所谓的 Oracle 攻击则在检测器已知的条件下试图制作一个非含水印产品。这类攻击所要考虑的是计算复杂度问题。

8.2.2 受限的水印操作

为进一步理解三类安全性攻击，下面分别用相应的三种场合来说明：为了保证安全性，哪些人可以进行哪些操作，而哪些人不可以进行哪些操作。考虑下面的三种场景^[5]。

场景 1：Alice 是一位广告商，在她将广告分发给 600 个广播站前，在每份广告中都嵌入一个水印。然后她可以使用水印检测器来对广播站进行监控和记录广告播出情况，并通过对比这些记录和她所收到的发票以发现虚假收费情况。假设 Bob 是这 600 家广播站中的一家，他想要在 Alice 所预定的时段播出自己的广告，然而却仍企图向 Alice 收取这段时间内费用。因此他秘密地将 Alice 的水印嵌入自己的广告时段，并以此来替代 Alice 原先预定的广告。这样，Alice 很容易被蒙蔽，通过检测到水印的存在而认为她的广告已被正常播出。

场景 2：Alice 拥有一套水印报告服务系统，并能在通过因特网传输的图像中加入水印用于所有者识别。Alice 也能广泛地向客户提供服务，帮助他们确认网上出现的含水印图像。她的客户便可凭借此信息识别出这些图像的任何非授权传播版本。而 Bob 制作了自己的网上“爬行检测器”（Crawler）用于检测由 Alice 所嵌入的水印，并能够提供更为廉价而具有竞争力的报告服务，从而将 Alice 的客户们吸引走。正是由于免去水印嵌入的成本，Bob 能够提供更为廉价的报告服务。

场景 3：Alice 拥有一间电影制作室，在电影发售之前，她在其中嵌入了用于复制控制的水印。她相信所有能够复制这些电影的记录设备都已包含了水印检测器并能够拒绝复制她的电影。Bob 是一个视频盗版者，能够设计出某种装置，用以去除存在于 Alice 电影中的用于复制保护的水印。使用此装置，Bob 可对 Alice 的电影进行非法复制。

在以上这些场景中，Bob 设法通过非授权操作来使 Alice 嵌入水印的用途失效。在场景 1 中，他进行了非授权嵌入的操作，即嵌入了本应只有 Alice 本人才能嵌入的水印。在场景 2 中，他进行了非授权检测操作，检测出了本应只有 Alice 本人才能检测出的水印。在场景 3 中，他进行了非授权去除操作，即去除了本不应去除的水印。这三类攻击中的每一个都能给安全水印设计者带来挑战。抵御每种攻击的重要性取决于具体的应用。通过指定谁有权利执行哪一项操作，可以对具体应用的安全要求进行分析。对一项给定的应用，可以将与之有关的整个人群划分为不同的几组，对每一组分配一个许可集

合。得到的操作表能够确定水印系统应具有的安全特性。

表 8.1 给出前述三种场景下对应操作表,“Y”表示这种操作必须被允许,“N”表示这种操作必须被禁止,“—”表示不论这种操作是否被允许,系统都要运行。注意到对一些条目而言,人们并不关心某个特定人群是否能够进行某些操作,因为往往这些操作并不影响系统的正常工作。然而,人们难以做到允许某一人群具有某一操作权限的同时剥夺他们进行其他操作的权限。例如,人们经常会问,是否一个水印系统可设计为在允许某个人嵌入水印的同时又能防止他进行水印去除操作?如果答案是不可能,人们就应该在复制控制的应用中拒绝给予公众嵌入水印的权限,将表 8.1 左下角的“—”改为“N”。

表 8.1 三种应用场景下的操作列表

应用场合	人 群	嵌 入	检 测	去 除
广播监控	广告商	Y	Y	—
	广播商	N	N	—
	公众	N	N	—
网页报告	标记服务	Y	Y	—
	报告服务	—	Y	—
	公众	N	N	N
复制控制	内容提供者	Y	Y	—
	公众	—	Y	N

为了进一步加深对前述“公有水印”和“私有水印”的理解,表 8.2 给出了公有水印和私有水印许可操作表。

表 8.2 两类水印应用(私有水印与公有水印)的操作表

应用类型	人 群	嵌 入	检 测	去 除
私有水印	受信任者	Y	Y	—
	公众	N	N	N
公有水印	受信任者	Y	Y	—
	公众	N	Y	N

需要指出,上面的两个表只是给出了一个何人被许可进行何类操作的大致说明,其中三列不同类别的操作下可能还有许多变型。即使对手难以完全进行某项未被授权的操作,他也可能会完成这项操作中的某一部分。例如,水印检测通常包括相分离的两个步骤:检测水印的存在性,以及对嵌入信息进行解码。在一些系统中,对手能够较容易地检测到水印的存在,但很难对消息进行解码。对一些应用而言,人们也关注这种部分攻击。因此,人们有必要将这三大类非授权行为按照更为具体的攻击类别进行区分。

8.2.3 关于对手的假设

在衡量水印技术能否满足表 8.1 和表 8.2 中的要求时,首先应对对手的能力作一些假设。他到底对水印算法有多少了解?他使用什么工具达到目的?例如,假设他想要去除某个水印,他是否清楚印记是如何嵌入的?他也有实验用的检测器吗?

1. 攻击者一无所知

最简单的假设便是对手对算法一无所知,也没有任何特殊工具(如水印检测器)。在

这种条件下，对手必须依赖于绝大多数广为人知的水印算法弱点。例如，假设对手认为作品已嵌入水印，想要将其去除。他可能会尝试使用其所知的多种不同的引入失真的方式来掩盖水印的存在，如去噪滤波、大压缩比的压缩、微小几何变换或时域失真。这些也是 Stirmark 程序所使用的基本方法，往往能使水印难以检测出来。

2. 攻击者拥有多于一幅的含水印作品

有时对手可以获得多份含水印作品。对手常利用这个条件来去除水印，即使他对算法一无所知。依靠多件作品的攻击称作共谋攻击。共谋攻击主要有两类：① 对手得到含有相同水印的多件不同作品，通过对其进行研究，得知算法的原理。② 对手获得相同作品的数件复制，每件复制嵌有不同的水印。这种情况下，对手可以通过结合不同的复制得到一个同原始作品极为接近的版本。

3. 攻击者知道算法

对于安全性要求高的系统而言，对手对水印算法一无所知的假设并不安全。对算法完全保密通常很困难。另外，若算法保密，则只会有少数研究者能够对其进行研究。这意味着系统的安全缺陷很可能直到系统安装时才会发现。因此加密研究人员坚持认为，应假定除一个或数个密钥之外，对手知道算法全部，这通常称作 Kerckhoff 假定。加密研究者不仅假定对手知道算法全部，他们还经常会公开发表这些算法以保证别人知道这些算法。这样可让同行研究这些算法，尽量找出算法的缺陷。知道全部算法的对手可在检测策略上寻找水印算法的弱点。例如，他可以找出特定的失真处理方法使检测器无法进行补偿，从而成功地实施掩盖攻击。另外，水印算法可以帮助对手确定描述特定水印算法的秘密。例如，一个生成重叠水印模式的算法（如基于分块的算法），已知水印抽取过程对重叠部分作平均，对手便可试着确定重叠部分的尺寸，生成参考模式的一个估计值。一旦对手得知这个秘密，他便可进行水印的非授权嵌入、非授权检测和非授权去除。

4. 攻击者拥有检测器

在前面所讨论的各种情况下，都假定对手可能得知算法的全部，但他并没有任何特殊工具可以利用。然而，如果应用场合规定了对手必须具有某些操作权限的话，就必须假定他具有这些操作所需的工具。其中研究最多的情况是允许对手进行水印检测，但并不允许去除水印。在这种情况下，必须假定对手拥有水印检测器。即使对手对水印算法一无所知，得到一个水印检测器对他而言也能帮助其对水印进行攻击。通常认为检测器对手而言是一个“黑箱”，送入一件经过处理的作品，可以判断其是否在检测范围之内。通过对作品作迭代更改，并在每次更改后进行测试，有可能得知很多检测算法的工作信息。如果对手拥有一个检测器，并且知道其如何运行，这个安全问题就变得更为严重了。从本质上说，他既然知道了检测算法，也就知道了检测特定印记所需要的密钥。目前，没有水印算法能够防止这类非授权去除攻击。然而，一些研究者也试图用非对称密钥的原理来解决这一问题。

8.3 非授权去除攻击

8.3.1 引言

非授权去除（Unauthorized Removal）是指通过攻击使作品中的水印无法检测到。一般说来，如果水印变得不可检测，便可认为水印已被去除。最极端的水印去除的例子是

作品复原为原始未嵌入水印时的形式。在所有需要防止非授权去除的应用中,防止对手恢复出原始作品都是必要的。然而很难想象出一个例子,仅仅具有这种防护措施就可以让系统具有足够的安全性。在绝大多数应用中,对手都试图修改含水印作品以使其同原始作品极为相似,从而使水印检测器无效。原始作品是许多符合此特性的作品中的一个。因此,使对手不能够复原出原始作品只是使水印具有防止非授权去除攻击能力的一小步。对手可以对逼真度进行约束,以使攻击所引入的失真具有最小的视觉影响。这类似于嵌入过程中用到的逼真度限制,但对手使用的视觉模型可能不同于嵌入过程中所使用的视觉模型。而且对手对逼真度要求通常不如嵌入过程严格。

在合乎对手要求的这些作品中,已不存在水印的那些作品不同于水印仍存在、只是需要比现有检测器更先进的检测器才能检测出来的那些作品。前者对应去除攻击,后者对应掩盖攻击。例如,考虑一个图像水印系统,其检测器无法检测出经过轻微旋转的图像中的水印。这样,对手便可以通过对图像作微小旋转来“去除”图像中的水印。然而,可以假设存在一个更为智能的检测器,采用了某种注册和搜索技术,能够作逆向旋转而检测出水印。这就是掩盖攻击的例子。与之相对应的是去除攻击,对手设法估计出嵌入水印的模式并将其从作品中去除。通过这种攻击能够得到同原始作品极为相近的一个版本(虽然没必要是原始作品的完全重构版本)。

直觉上,去除攻击和掩盖攻击的差别很明显。然而在媒体空间中考虑该问题时,区别不再明显。旋转攻击和估计-相减攻击都是将检测区域(即含水印作品)的点移向检测区域之外(即不含水印的作品)。使用能够进行几何搜索操作的检测器只是将检测区域变成包含受旋转攻击的作品。如果能够通过设计更为先进的检测器来抵抗旋转攻击,为什么不能够采取类似方法来抵抗所有其他攻击?显然,能够完全恢复原始作品的攻击不可能通过改进检测器进行抵抗。通过检查受攻击作品同不含水印作品在媒体空间中的分布关系,可辨识出其他一些同样不能用此方法进行抵抗的攻击。在旋转攻击例子中,被攻击作品仍然具有嵌入过程引入的一些特性。也就是说,它们不太可能呈现不含水印作品的分布形式。因此,将检测区域变为包含被攻击作品区域的方法并不能明显地增高误检率。另一方面,经过估计-相减式的攻击得到的作品中很可能出现未嵌入水印作品。如果检测区域包含了这些受攻击作品区域的话,误检率将大大提高而达到不可接受的程度。基于上述,可将去除攻击定义为将含水印作品移向媒体空间中很可能出现不含水印作品区域的攻击形式,而掩盖攻击可定义为将含水印作品移向媒体空间中不太可能出现不含水印作品区域的攻击形式。

在一些被攻击后能方便改进检测器的场合,去除攻击和掩盖攻击的差别显得十分重要。例如,在交易跟踪应用中,如果对手想通过掩盖攻击破坏系统以使所有者的水印检测器无法检测出盗版作品中的水印,所有者会有机会升级其检测器从而检测出失真的水印。另一方面,如果对手能够进行去除攻击,那么所有者的检测器即使拥有升级能力也难以检测出水印。因而在有可能升级检测器的应用场合,掩盖攻击不如去除攻击严重。

8.3.2 去除攻击

通过对内容降质而去除水印是所有攻击中最笨和最直接的方法,也是文献中研究最多的攻击方法。当水印方案的设计者试图达到工程上的鲁棒性要求时,他们必须考虑到各类攻击方法。设计出的方案必须能抵抗常规的攻击,像压缩、裁剪、模糊化、甚至打印和重扫描,或者还希望方案对于预料不到的操作也具有鲁棒性。但不幸的是,攻击者

拥有图像处理操作的庞大武器库来进行攻击，对任何方案的简要分析经常能找到删除水印的一些简单操作方法。下面介绍常见的几类去除攻击。

1. 线性滤波（去噪）攻击

若想水印具有抵抗非授权去除的安全性，就必须使其对任何保持作品逼真度的操作都具有鲁棒性。这个操作可能是正常操作，对应鲁棒性攻击方式；然而，它也可能是一种恶意操作，对应安全性攻击方式。任何能够保持作品逼真度的操作都可被对手用来去除水印。

线性滤波也可能被对手用来达到水印去除的目的。例如，如果水印的主要能量集中在高频段的话，就可利用低通滤波进行攻击。另外，任何基于“类似噪声”添加方式的水印系统都会对噪声去除技术极为敏感。

对某些类型的水印系统而言，文献[133]表明**维纳滤波**（Wiener Filtering）是最优的线性滤波/噪声去除攻击方式。如果满足以下条件，维纳滤波可能是最具破坏性的线性时不变过程：① 添加模式独立于隐藏作品；② 作品和水印都服从零均值的正态分布；③ 线性相关作为检测统计手段。此外，文献[134]提出了一种使失真最小的攻击方法。该方法把攻击模型化为线性时不变滤波和加性噪声的组合

$$\hat{\mathbf{x}} = \mathbf{x}^w * \mathbf{h} + \mathbf{n} \quad (8.5)$$

其中， $\hat{\mathbf{x}}$ 为攻击后的含水印图像， \mathbf{h} 为滤波器系数向量， $*$ 为卷积算符， \mathbf{n} 为噪声。在此基础上，还定义了攻击失真。攻击时通过一定的方法使该攻击失真最小。

2. 压缩

JPEG 压缩是当前静止图像应用最为广泛的压缩算法。当人们准备将图像发布到网上时，图像要被调整大小、压缩以适应版面设计和带宽需求。不幸的是，有损压缩会删除一些对可视性影响微小的高频分量，而只保持低频分量。这就影响了某些数字图像水印方案，这些方案的原理就是将信息嵌入到高频部分以减少失真。因此，有人建议，水印应该放在图像对感观影响重要的成分中，而不管它所可能引人的失真，但这样可能会留下可视的人为痕迹。此外，MPEG 压缩或重压缩所引入的量化噪声可能会使嵌在视频中的水印消失。

3. 共谋攻击

共谋攻击（Collusion Attacks）也称共谋攻击或串谋攻击^[5]，它主要有两种情况。在第一种情况下，对手得到含有相同水印的多件不同作品，通过对其进行研究，得知算法的原理。最简单的例子是对手对几件不同的作品做平均。如果所有的作品均被添加了相同的参考模式，这种平均将会得到近似的模式。将平均后所得到的模式从作品中减去，即可去除水印。设同一个水印 \mathbf{w} 被嵌到 Cox 和 Linnartz 所描述的视频序列的一组图像 $\{\mathbf{x}_i\}_{i=1}^n$ 上^[135]，若 f 是特征提取函数（描述水印应加到哪里）。在这个特征域中， n 个含水印帧加起来就是 $n\mathbf{w} + \sum_{i=1}^n f(\mathbf{x}_i)$ 。当 n 值很大时，它的期望值是 $n\mathbf{w}$ 。这种情况的一种变型

如下：如果相同的水印嵌入在一件作品的好几个部分，对手便可将这几个部分看作独立的作品，通过共谋攻击来识别出模式。这种攻击已被成功地用在一个音频水印系统 SDMI（Security Digital Music Initiative）之中^[136]。

在另一种共谋攻击中，对手获得相同作品的数件复制，每件复制嵌有不同的水印。这种情况下，对手可通过结合不同的复制得到一个同原始作品极为接近的版本。最简单

的结合方式就是对几份复制做平均, 这样将不同的印记混同以减小其幅度。图像的发行商为每一位顾客嵌入一个唯一标识号以跟踪侵权者。这样, 图像是相同的而水印是互不相同的, 所以含水印图像合计为 $nf(\mathbf{x}) + \sum_{i=1}^n \mathbf{w}_i$, 当 n 很大时, 它的期望值为 $nf(\mathbf{x})$ 。文献[137]引入非线性共谋攻击的概念, 在攻击时采用多种非线性函数 (而不是简单平均), 如最大值、最小值、中值、最大值和最小值的平均值、最大值加最小值减中值等。

4. 块替换攻击

文献[138]中提出了一种块替换攻击方法。该攻击方法的核心思想是使用邻块的插值形式来代替当前图像块。以前人们通常用这种替代算法来恢复丢失的块或者进行低比特率编码。因为简单的插值 (比如双线性插值) 通常会产生具有模糊边缘和纹理的质量很差的图像, 所以更好的方法就是首先使用与丢失块相邻的块来推断边缘信息, 然后对丢失块进行插值给出边缘。现在把该方法用在攻击中, 需要保持处在图像边界的所有块不变, 而对所有的其他块利用临近块的插值形式所代替。首先把一个图像分成 4×4 或者 8×8 块。对除边界块外的每一块, 都能从相邻块中获得它的插值形式。如果插值形式接近于原始块, 那么就用它来替换原始块。很容易看出这样逐块地替换必然会毁坏最初嵌入每一块中的水印。实际上, 这种攻击方法可看作是一个非线性低通滤波器。

5. 频率模式拉普拉斯去除攻击 (Frequency Mode Laplacian Removal, FMLR)

文献[139]提出一种频率模式拉普拉斯去除攻击 (FMLR) 方法。首先介绍拉普拉斯去除攻击方法 (LR)。LR 攻击方法利用图像边缘检测中的拉普拉斯 3×3 卷积模板。被攻击的图像首先用拉普拉斯卷积模板进行处理。处理所得的图像称为“负拉普拉斯图像” (L_n), 该图像的很多成分反映出原始图像修改 (即水印嵌入) 的逆过程。对负拉普拉斯图像再应用一次拉普拉斯卷积模板, 得到正拉普拉斯图像 (L_p), 它包含很多水印信息成分。如果将含水印图像记为 \mathbf{x}^w , 攻击后的图像记为 $\hat{\mathbf{x}}$, 则原始的 LR 算子是

$$\hat{\mathbf{x}} = \mathbf{x}^w - \alpha(L_p - L_n) \quad (8.6)$$

这里, α 是攻击强度, 成功去除水印的 α 典型范围是 0.05~0.15。

FMLR 攻击算子以相似的方式工作, 但要在对上述三个图像进行 DCT 变换之后。首先将 \mathbf{x}^w , L_p 和 L_n 分成 8×8 像素块, 然后对各块分别作 DCT 变换。对于每一块, 进行下面的计算

$$\hat{\mathbf{x}}_f(i, j) = \mathbf{x}_f^w(i, j) - \alpha(L_{pf}^\gamma(i, j) - L_{nf}^\gamma(i, j)) \text{Mod}(i, j) \quad (8.7)$$

这里, $\hat{\mathbf{x}}_f$ 表示攻击后图像的变换域系数块, \mathbf{x}_f^w 表示含水印图像的变换域系数块。 L_{pf} 和 L_{nf} 分别是 L_p 和 L_n 的分块 DCT 变换域系数块, $0 \leq i, j \leq 7$ 。 α 是强度, γ 取 0.3 能得到最好的结果。 **Mod** 是含有 64 个值的数组, 它可对攻击算子的有效性进行精确地控制。

6. 盲模式匹配攻击 (Blind Pattern Matching, BPM)

文献[140]提出了一种盲模式匹配攻击 (BPM) 方法。BPM 攻击不受内容类型或特殊水印算法的限制。为了成功地发起攻击, 对手不必知道水印编解码器的细节。但对手需要减少块尺寸, 以致无法检测各块所含的水印信息。对于音频和视频, 这一要求不难。对于音频代表性的是用含 128-1024 变换系数的块, 对于视频代表性的是用 64×64 像素的位图进行模式匹配。该攻击的主要步骤如下。

(1) 信号分割。在这一步中, 含水印载体 \mathbf{x}^w 被分割为一组重叠块, 第 p 块的起始位置

索引为 B_p ，即起始于 $x^w(B_p)$ ，长度为 m 。设重叠率为 η ，则块数 $n = [(N - m) / (1 - \eta)]$ 。重叠率越高，BPM 攻击的搜索空间就越大，但要遵循下面的几点。

1) 连续块没有相似的感知特性，块重叠的上限以减少搜索空间为目标。

2) 对始于 $x^w(B_p)$ 和 $x^w(B_{p+1})$ 的两个连续块，始于 $x^w(a)$ ， $a = [B_p + B_{p+1}] / 2$ 的块在感知上与第 p 或 $p+1$ 块不相似。

(2) 相似函数。这是 BPM 攻击的核心函数。若输入第 p 块和第 q 块，返回两块的相似度 $\varphi(B_p, B_q) \geq 0$ 。当 $\varphi(B_p, B_q) = 0$ 时，代表两块相等。敌手可对许多不同的函数进行试验。在这里，相似度定义为平方欧几里德距离，即

$$\varphi(B_p, B_q) = \sum_{i=0}^{m-1} (x_{i+B_p}^w - y_{i+B_q}^w)^2 \quad (8.8)$$

(3) 模式匹配。这一步识别信号块两两之间的感觉相似性，得到一个对称的二值相似矩阵 \mathbf{S} ，其元素如下

$$S_{pq} = \begin{cases} 1, & \text{若 } m\alpha^2 \leq \varphi(B_p, B_q) \leq m\beta^2 \\ 0, & \text{否则} \end{cases} \quad (8.9)$$

这里， α^2 和 β^2 分别表示可替换的两块的最小和最大平均相似性的参数。因为用特别相似的另一块取代将不影响水印检测，所以这里需要下限。而上限用来确保高保真度。

(4) 块置换。在最后一步中，根据相似度进行块替换，获得了攻击后的信号 $\hat{\mathbf{x}}$ 。BPM 攻击的块尺寸是一个需要折衷的变量。很难发现大的相似块，因此很明显搜索应该用较小的块。另一方面，很难估计小块的视觉因子。另外，较小的块往往在原始块和替换块之间保持较高的相关性，这一现象减小了 BPM 对水印检测器可靠性的影响。最后，较小的块增加了需要置换的块数量，这样将极大地增加搜索时间。

7. 专门设计的攻击

上面介绍的几种攻击方法并不是针对某种水印算法专门设计的，具有一定的通用性，也就是说假设对手不知道水印算法。如果对手了解水印算法的细节，攻击者可以发明一种攻击，专门设计来删除某种水印。下面介绍几种专门设计的攻击。

(1) 针对扩频的非线性滤波

由于很多水印方案基于扩频技术，Langelaar 等^[14]描述了如下攻击思想：把一个含水印图像 \mathbf{x}^w 分成两部分， $\hat{\mathbf{x}}$ 和 $\hat{\mathbf{w}}$ ，这样一来估计图像 $\hat{\mathbf{x}}$ 就不再包含水印了。Langelaar 等通过实验发现， 3×3 中值滤波应用于 Bender 等提出的方法^[16]时能提供很好的分离结果，然后把它用于水印的粗略估计，得到 $\mathbf{x}^w - \text{med}_{3 \times 3}(\mathbf{x}^w)$ 。在从 \mathbf{x}^w 中减掉之前，这个估计信号还应该提炼一下，因为它仍包含边缘信息，一些值可能太大。后来他们建议把它通过一个高通滤波器 H 滤波，并将输出值调整在 $[-2, 2]$ 范围之内，即估计的原始图像和水印表示如下

$$\begin{cases} \hat{\mathbf{x}} = \mathbf{x}^w - \hat{\mathbf{w}} \\ \hat{\mathbf{w}} = aH_{3 \times 3}(\mathbf{x}^w - \text{med}_{3 \times 3}(\mathbf{x}^w))|_{[-2, 2]} \end{cases} \quad (8.10)$$

其中， a 是实验中确定的放大系数。

(2) 对回声隐藏的攻击

对付回声隐藏的攻击显然是检测回声，然后通过简单地逆转一下卷积公式来删除它，问题关键是在于不了解原始载体和回声参数的情况下检测回声。回声隐藏系统是基

于倒谱分析的, 攻击者可用相同的检测函数, 但需要把它与强力搜索结合在一起。

以下倒谱分析方法由 Bogert 等提出^[142]。设信号 $y(t)$ 中包含一个简单的单回声, 即

$$y(t) = x(t) + \alpha x(t - \Delta t) \quad (8.11)$$

如果定义 Φ_{xx} 为 x 的功率谱, 那么

$$\Phi_{yy}(f) = \Phi_{xx}(f)(1 + 2\alpha \cos(2\pi f \Delta t + \alpha^2)) \quad (8.12)$$

其对数近似于

$$\log \Phi_{yy}(f) \approx \log \Phi_{xx}(f) + 2\alpha \cos(2\pi f \Delta t) \quad (8.13)$$

这是频率 f 的函数, 功率谱增加了“倒频” Δt , 就是 $\cos(2\pi f \Delta t)$ 的频率。后一项的自协方差峰值出现在“倒频” Δt 上。如果用如下倒频的修改版, 则结果会好些

$$C \circ \Phi \circ \ln \circ \Phi \quad (8.14)$$

其中, Φ 是功率谱密度函数, \circ 是组合运算, C 是如下自协方差函数

$$C(x) = E((x - \bar{x})(x - \bar{x})^*) \quad (8.15)$$

对一些像音乐这样的随机信号做实验, 结果表明当一个人工回声加到信号上时, 上述估计方法能返回相当准确的时延估计值。在检测函数中, 只有回声延迟在 0.5 与 3 毫秒之间是可以的, 低于 0.5 毫秒, 函数就不能正确工作, 而高于 3 毫秒, 回声就能听出来了。顺便提一句, 在原始回声隐藏系统中, 回声延时选择在 0.5 和 2 毫秒之间, 具有最好相关振幅的回声是在 0.8 毫秒附近。

(3) 直方图攻击

在一些情况下, 待加水印的图像具有特定的性质, 这些性质可帮助恶意攻击者获取水印本身的信息。例如, 某些图像只有少量不同颜色, 如卡通图片, 它在颜色直方图上有明显的峰值。直方图攻击利用这些性质来恢复简单的扩频水印。下面, 在灰度图像情况下阐述这种攻击。

一种简单的扩频数字水印算法就是对每一个像素随机地加上或减掉一个固定值 d 。于是每一个像素值有 50% 的机会增加或减少。设 n_k 是灰度值为 k 的像素个数。假设对特定灰度值 k_0 , 与它相邻的第 d 个临近灰度不会出现, 故 $n_{k_0-d} = n_{k_0+d} = 0$ 。因此, 加水印后, 期望的像素个数为: $\hat{n}_{k_0-d} = \hat{n}_{k_0+d} = n_{k_0} / 2$ 而 $\hat{n}_{k_0} = 0$ 。所以, 用一组类似的等式, 在某些特定情况下, 可恢复出原始直方图的分布和嵌入水印的值。

(4) 补位攻击

文献[143]提出一种针对 HCH 嵌入方案 (由 Hwang, Chang 和 Hwang 三人提出)^[144] 的补位攻击方法。该方法对含水印图像的一些位取补, 也就是说, 若原来的位是 0, 那么用 1 来代替这个 0, 否则相反。实验表明, 对用 HCH 嵌入方案得到的含水印图像所有像素的最不重要的 2, 3, 4 位取补后, 所嵌入的水印可能会被彻底毁坏, 但相应信噪比很低 (28.53dB)。为了在应用补位攻击后仍保持较高的信噪比, 文献[143]建议仅仅对不会引起含水印图像质量重要影响的像素取补。换句话说, 在对一些位取补时所做的变换不会引起图像质量严重退化的情况下, 这一像素值才变换成为另一像素值。

8.3.3 掩盖攻击

使水印嵌入算法无效, 不一定必须删除含水印作品中的水印, 而可通过对内容进行修改, 使得检测器找不到有效水印。这类攻击也叫**表达攻击** (Presentation Attack), 主要包括**几何攻击** (Geometric Attacks)、**拼凑攻击** (Scrambling Attacks)、利用检测器的攻击

和加噪攻击四种。前面两种都可看成是**同步性攻击**（Synchronization Attacks）。所谓同步性攻击是指对手通过打乱同步来掩盖水印，大多数数字水印技术对同步攻击都很敏感。

1. 几何攻击

简单的同步性攻击包括音频和视频的延时和时间缩放，图像和视频的旋转、缩放和平移，即所谓的仿射变换攻击^[144]。更为复杂的失真包括音频中的**音调保持缩放**（Pitch-Preserving Scaling）和**样值去除**（Sample Removal），图像中的**剪切**（Shearing）、**水平镜像**（Horizontal Reflection）及**行列去除**（Column or Line Removal）等。甚至可能出现更为复杂的失真操作，如图像的非线性弯曲。StirMark 程序可模拟这些失真操作。下面介绍两种常见的几何攻击：一种针对图像；一种针对音频。

（1）针对图像的几何攻击

虽然很多水印系统能抵抗基本的处理，即那些用标准工具就能轻松实现的处理，但它们往往不能应付这些处理的合成或微小随机几何变形。

StirMark 是一种水印系统的测试工具，它可对图像进行几何变形。若 A 、 B 、 C 、 D 为图像的四个角，则图像内的点 M 可表示为

$$M = \alpha(\beta A + (1 - \beta)D) + (1 - \alpha)(\beta B + (1 - \beta)C) \quad (8.16)$$

这里， $0 \leq \alpha, \beta \leq 1$ ，是 M 相对于四个角的坐标。StirMark 中的随机双线性失真是通过用一个很小的随机数从两个方向上改变角坐标来实现的。保持 (α, β) 不变，但用四个新角坐标计算即可得到新的 M 值。这种变换是可逆的。该失真不能从根本上删除水印，但它可阻止一些系统检测或恢复出水印。

除上面的随机双线性失真，还可对图像中的每个像素进行微小偏离，而边缘像素几乎不做处理。在实际实现时，这种操作可用简单的正弦波来模拟。若 (u, v) 为某像素坐标， $0 \leq u \leq U$ 和 $0 \leq v \leq V$ ，则变化后的像素坐标 (u', v') 为

$$u' = u + \lambda \sin(\pi v / V), \quad v' = v + \lambda \sin(\pi u / U) \quad (8.17)$$

在此基础上作一个高频位移，令

$$\delta = \lambda \sin(\omega_u u) \sin(\omega_v v) (1 + n(u, v)) \quad (8.18)$$

其中， n 是随机序列，则

$$u'' = u' + \delta, \quad v'' = v' + \delta \quad (8.19)$$

所有这些变形与适度 JPEG 压缩结合在一起，也不能从本质上删除水印，但它们能阻止检测器找到水印，能使检测器失去同步。这就说明在数字水印应用中，真正的问题不只是加入水印，还必须能把它辨别出来。这些微小的随机几何变形还可用于视频。但是，这只是视频的一个方面，更好的攻击应该同时考虑视频的时间轴，否则视频播放会出现抖动。

（2）针对音频信号的跳跃攻击（Jittering Attacks）

跳跃攻击^[145]主要用于对音频水印系统进行攻击，其一般实现方法如下：在音频信号上加入一个跳跃信号，即首先将信号数据以 500 个采样点为一个单位进行分块，然后在每个数据块中随机复制或删除一个采样点，来得到 501 或 499 个采样点的数据块，接着再将数据块按原来顺序重新组合起来。实验结果表明，即使对古典音乐信号数据这种改变也几乎感觉不到，但是却可以非常有效地阻止水印信号的检测定位，以达到难以提取水印信号的目的。类似方法也可用来攻击图像数据的数字水印系统，其实现方法也非常简单，即只要随机地删除一定数量的像素列，然后用另外的像素列补齐即可，该方法虽然简单，但仍然能有效破坏水印信号存在的检验。

2. 拼凑攻击

拼凑攻击 (Scrambling Attacks) 实际上是一种系统级别的攻击^[5], 它在作品样本呈交给水印检测器前进行拼凑, 经过检测后进行解拼凑。拼凑的方式可以是简单的排列或更为复杂的伪随机拼凑。唯一的要求是拼凑过程必须是可逆或近似可逆的。近似可逆或有损的拼凑攻击应得到感官上同含水印作品相近的作品。

马赛克攻击 (Mosaic Attack) 是广为人知的一种拼凑攻击^[5], 它把图像分割为很多小矩形块, 每一块都小得无法进行可靠的水印检测。这些图像片段再以边沿相接的方式组合显示出来。这些小图像块的组合结果在感官上同分割前的图像完全相同。这种技术可用于网页中, 以顺利通过**网页过滤检测器 (Web-Crawling Detector)**。可以看出, 拼凑过程实际上是原始图像分解为子图像的过程, 而解拼凑过程则由网页浏览器本身完成。很明显, 图像越大, 越容易隐藏一定数量的比特信息。反过来也一样, 这就是马赛克攻击的基础, 即一个图像块可小到不能检测到水印。这一点在表达攻击中非常重要, 它有普遍的适用性。

马赛克攻击对对手而言十分方便, 因为绝大多数网页浏览器会正确地将图像“解拼凑”。更一般的拼凑攻击需要盗版作品的接收端具有解拼凑设备或程序。例如, 为了绕过视频录制设备的复制控制系统, 消费者可能会购买并不昂贵的拼凑器和解拼凑器。通过对视频录制设备的输入进行拼凑, 对手可以使水印变得不可检测, 录制设备便能够允许继续复制。在回放前, 录制的拼凑视频经过解拼凑设备便可重新观赏。虽然许多国家已立法禁止销售这类企图绕开版权法的装置, 但拼凑设备确实还有合法的用途, 如帮助成人防止儿童观看不适宜的内容。

3. 利用检测器的攻击

在很多应用领域, 攻击者可访问到水印检测器。这种检测器可能是配有常用图像处理包的软件, 或者是像 DVD 那样嵌入到电器中的电子电路。即使攻击者不知道水印嵌入方法, 他仍然可使用检测器返回的信息来删除水印, 对图像作微小变动直到检测器什么都找不出来。下面介绍几种这样的攻击方法。

(1) 敏感分析攻击 (Sensitivity Analysis Attacks)

敏感分析攻击也叫 Oracle 攻击, 是一种用于水印非授权去除的技术, 在攻击时假设对手已经拥有一个“黑盒”检测器。敏感分析攻击使用检测器来估计出从含水印作品到检测边缘区的最短路径的方向。我们假设这个最短路径可以由检测区域的法线 (Normal) 来近似, 而这个法线在检测区域的很多部分都可看作常数。本章所述的相关检测方法满足这些条件。这种攻击可以分作三步。第一步要找出一件靠近检测区域边界的作品 A, 这件作品在感官上并不需要同含水印作品相似 (如果相似的话, 也没有继续处理的必要了)。通过改动含水印作品可使用数种方法逼近检测区域的边界。主要的方法有降低含水印作品的幅度 (对比度或绝对值), 用作品的均值替代样值, 或者将含水印作品同不含水印的作品作线性组合。在这三种情况下, 失真都可逐渐增加直到水印不能再被检测到。第二步便是对作品 A 检测区域表面的法线方向进行近似, 这可通过迭代方法做到这一点。一旦法线方向被估计出来, 第三步即是将此法线缩放并从水印作品中减去。目前, 已有两种方法提出用来估计 N 维法线矢量。在文献[146]中, 通过考虑 N 个独立的修正矢量来找出这个法线矢量。其中每一个矢量都通过增加幅度分别在作品 A 上添加或减去, 直至不再能够检测到水印为止。此时检测区域表面的法线可用 N 个修正矢量的加权和来近似, 其中每一个矢量均可用其缩放因子加权, 缩放程度越大, 权值越小。在文

献[147]中, 使用迭代的方法估计作品 A 检测区域表面的法线。每一次迭代都在作品 A 上添加一随机矢量, 并记录检测结果。如果得到的作品具有肯定的检测结果, 则这个随机矢量同法线的估计值相加。如果得到的作品不含水印, 这个随机矢量则从法线的估计值中减去。

(2) 特定的 Oracle 攻击

上述 Oracle 攻击只能对付公开水印, 但即使是纯秘密水印, 特定的 Oracle 攻击也可能成功。这种攻击唯一的需求就是攻击者要有水印嵌入和检测的算法。攻击者用和版权拥有者相同的方法在图像中一次或多次嵌入自己的水印。攻击者用自己的水印作为秘密水印的随机强度指示器, 然后用普通的 Oracle 进行攻击, 直到所有新嵌入的水印都被删除掉。因为原始水印随着图像随机改变而变弱, 当加入其他水印时, 它也会变弱。所以可以假设, 新加水印的强度为原始水印强度提供一个上限。这样一来, 当所有新水印都被删除掉时, 原始秘密水印就会以很高的概率被删除掉。

(3) 梯度下降攻击

梯度下降攻击^[5]与敏感分析攻击不同, 对手需要拥有一个能够给出实际检测值而非最后的二值结果的检测器。逐渐改动作品时, 对手利用检测值的变动可以估计出含水印作品检测统计值的梯度。这里设最陡梯度下降方向就是越出检测区域的最短路径方向。给定一件含水印作品, 可使用任意的搜索策略来确定最速下降的局部梯度。作品可沿此路径方向做少量变动, 不断迭代这个过程直至得到的作品恰好落于检测区边界之外。成功进行这种攻击的前提是局部梯度必须指向到达边界的最短路径方向。这对许多统计检测方法而言都适用, 包括线性相关和归一化相关。为防止这种攻击, 检测区域内的统计方法不应表现为向边界处单调递减。相反, 它应包含许多局部最小点, 使得局部梯度方向无法帮助确定越出检测区域的最小路径方向。

4. 加噪

除了上面三种攻击方法, 人们还可能考虑运用在含水印图像中加入一定量的噪声来实施攻击。但是, 文献[135]指出人们一个普遍的误解是: 通过加入与水印相同幅度的随机噪声可去除水印。实际上, 相关检测器对这种随机噪声攻击很鲁棒。设攻击后的图像为 $\hat{\mathbf{x}} = \mathbf{x} + \mathbf{w} + \mathbf{\varepsilon}$, 相关检测器计算内积

$$\hat{\mathbf{x}} \cdot \mathbf{w} = \mathbf{x} \cdot \mathbf{w} + \mathbf{w} \cdot \mathbf{w} + \mathbf{\varepsilon} \cdot \mathbf{w} \quad (8.20)$$

如果水印算法能保证 $\mathbf{w} \cdot \mathbf{w}$ 远远超过其他两项, 则加入的噪声根本不起作用。因此, 在实际攻击时, 噪声并不是主要威胁, 除非噪声比图像 \mathbf{x} 大或者噪声与水印相关。

8.3.4 对策

虽然水印和加密之间的相似性对防止非授权嵌入和解码十分有用, 但在非授权去除的情况下, 它的应用并不明显。这是因为在主流的加密问题研究中并不存在同非授权去除类似的问题。严格地讲, 非授权去除完全是传输层的问题: 对手会阻止消息到达接收地点。当对手未被授权进行嵌入和解码操作时, 人们可使用扩频技术而非加密技术来防止非授权去除。此时问题变得类似于安全军事通信, 其中扩频技术可以保证消息的可靠传达。利用扩频通信, 可以在远大于最小所需的带宽范围内传送一个信号。信号的扩频分布由密钥控制, 接收端也必须拥有这个密钥以检测信号。扩频通信很难被阻塞或去除, 非授权检测的概率也非常小。

通过建立类似于非对称密钥加密系统的方式保证水印处理系统的安全是正在研究的

一个方向。即人们可以让水印嵌入器和检测器所使用的密钥不相同。前提是即使得知检测密钥，对手也很难将水印去除。已有数种使用非对称密钥的水印系统提出，它们或者在嵌入和检测阶段使用不同的密钥，或者只在嵌入过程中使用密钥而对检测过程不设置密钥。虽然这些方法具有同非对称加密类似的非对称性，但尚未见到这些方法用于防止水印去除操作。事实上，这些方法中的绝大多数都易受到诸如敏感性分析之类的攻击。

值得注意的是，对于知道检测算法和密钥的对手的非授权去除攻击，非对称密钥水印系统可能并非唯一的解决方法。例如，也许可以构建一个检测区域同嵌入区域相异的安全系统。嵌入区域应较为简单，以使人们容易在其中找到一个与区域外的某一点很接近的点。而检测区域应较为复杂，以使得无论怎样描述这个区域，都很难在其外找到一点与其内的某一点很接近。如果检测区域包含嵌入区域，水印的嵌入和检测都会变得很容易，但要去除水印却十分困难，即使在完全知道嵌入和检测密钥的情况下也是如此。如果这样的系统可被实现，系统无须保证密钥的安全也能保证对去除操作的安全性。

下面介绍针对几种典型非授权去除的对策。

1. 针对维纳滤波

文献[133]指出，通过选择同原始作品功率谱相似的添加模式的功率谱，水印抵抗维纳滤波的安全性可以达到最大，即

$$|\mathbf{w}_a|^2 = \frac{\delta_{w_a}^2}{\delta_{x^w}^2} |\mathbf{x}^w|^2 \quad (8.21)$$

其中 $|\mathbf{x}^w|^2$ 是含水印作品的功率谱， $|\mathbf{w}_a|^2$ 是添加模式的功率谱，而 $\delta_{w_a}^2$ 和 $\delta_{x^w}^2$ 是添加模式和含水印作品分布的方差。这是一个感知上的结果，它表明当水印“看起来很像”含水印作品时，对手要想分离二者是很困难的。

2. 针对共谋攻击

对共谋攻击很显然的一种对策是嵌入多个水印，并让它们在图像中相互独立。这个问题，在数字指纹技术中已经详细介绍。Boeh 和 Shaw^[50]提出了**共谋安全编码**（Collusion-Secure Codes）的设计问题（即能够抵御这种攻击的编码）。设想一件作品被分发给了 n 个人，每件拷贝中嵌有不同的编码字串。如果在最多 c 个人共谋时，处理后的一件作品含有足够信息识别出至少一个共谋者的可能性很高，那么这种编码称作 c -安全（ c -secure）。如果所有共谋者的编码字串的某一部分均相同，那么在对这些拷贝进行对比时，共谋者不会知道他们仍需要破坏这一部分，因此人们假设这一部分不会遭受攻击。如果未受影响的这一部分编码字串携带有足够信息，人们便能够至少指认出其中一个共谋者。

3. 针对几何攻击

有人也许会试图通过预测盗版者所可能使用的变换来增加水印系统的鲁棒性，而如果这些变换是难以预测的，这种做法就会很难实现。O'Ruanidh 和 Perera 建议使用 Fourier-Mellin 变换来解决旋转和缩放问题^[148]。人们还可以利用几何变形在局部几乎都是线性的这个事实（等价于平移和旋转），用基于块的检测算法。当原始图像可以得到时，人们足可以近似地对随机变形进行推断。

4. 针对 Oracle 攻击

一个可能的 Oracle 攻击对策是对检测过程随机化。假设检验不是使用一个门限，而是使用两个门限。在两个门限值之间，检测器得到一个随机答案，在第一个门限值之下

表示“存在”，第二个门限值之上表示“不存在”，这种随机化决策规则将大大阻止依赖微小改变水印图像的攻击。当然，也可以把解码过程复杂化。但是，在没有防篡改硬件的情况下，无论哪种方法都不是非常令人满意。需要公共验证的水印应用（如 DVD）好像注定要在防篡改技术的约束下进行。最后可能的对策就是依赖于数字水印的意义了。如果水印存在，意味着可以通知记录设备有人试图非法复制，那么上述设备就有理由不发挥作用或者对复制企图发一个悄悄的警告。如果有五次非法复制之后，攻击者就必须带着 DVD 机去商店重新启动。

8.4 非授权嵌入攻击

8.4.1 引言

非授权嵌入（Unauthorized Embedding）是指在作品中嵌入本不该含有的非法水印信息。最为彻底的一类非授权嵌入攻击是：对手伪造和嵌入一个自己的原始消息。例如，假设 Alice 开展一项场景 2（见 8.2.2 小节）中的业务，不同的是她只向用户收取嵌入所有者 ID 的费用，而免费发放其网络监控软件。如果 Bob 想要无偿使用这个网络监控软件，他会设法嵌入自己的身份标记。为此他必须编制一个新的消息以声明某件作品为他所有，并构造一个嵌入工具。另一种不太彻底的非授权嵌入方式是预先编制好一个合法消息（而不用编制新的消息），并将此消息非法地嵌入到一件作品中。例如在场景 1 中，Bob 企图在广播前，将 Alice 的 ID 嵌入到他的广告中。如果 Alice 的水印系统设计欠佳，Bob 就可能辨认出 Alice 在广告中对自己 ID 信息编码的参考模式。这样，Bob 只需将这些参考模式直接复制到他自己的广告中，往往并不需要知道此消息是如何被编码的。

本节将非授权嵌入攻击方法分为四类。第一类是复制攻击，第二类是多重嵌入攻击，第三类是协议攻击，第四类是针对脆弱水印的非授权嵌入攻击。其中，前三类均针对鲁棒水印算法，复制攻击是把估计出的他人水印模式嵌入到其他作品中，多重嵌入攻击是在别人的含水印作品中再一次嵌入自己的水印，协议攻击实际上并没有真正嵌入水印，而是伪造出自己的原始图像和水印使版权解释不清（实际上，这类攻击可归入系统攻击）。而针对脆弱水印的非授权攻击主要包括三种：① 生日攻击；② 复制攻击；③ 二次拍摄攻击。下面将分别详细介绍这四类攻击。

8.4.2 复制攻击

复制攻击（Copy Attack）是指将水印从一件作品复制到另一件作品。复制攻击一词由文献[149]提出，该文还介绍了一种特定的实现方法。假设存在一件合法的含水印作品 \mathbf{x}_1^w ，以及一件不含水印的其他作品 \mathbf{x}_2 。首先对 \mathbf{x}_1^w 使用水印去除攻击得到原始作品的近似版本 \mathbf{x}' 。这一步可使用非线性降噪滤波器或其他任何能够估计原始作品的方法。然后，通过从含水印作品中减去所估计的原始作品来估计所嵌入的水印模式

$$\tilde{\mathbf{w}}_a = \mathbf{x}_1^w - \mathbf{x}' \quad (8.22)$$

最后，将估计得到的水印模式添加到不含水印的作品 \mathbf{x}_2 以得到其含水印版本

$$\mathbf{x}_2^w = \mathbf{x}_2 + \tilde{\mathbf{w}}_a \quad (8.23)$$

应用表明，该方法对许多商业图像处理系统能进行有效地攻击。

文献[149]的方法需要得到原始作品 \mathbf{x}_1 的近似版本 \mathbf{x}' 。因此，如果能保证无法让对手

得到这样的近似版本,便可阻止这种攻击。然而,这与水印算法有关。对于有些水印算法,在进行这种攻击时可能无须得到作品 x_1 的近似版本 x' 。一个简单的例子就是针对将水印嵌在作品最低有效位上的水印算法。由于原始作品的 LSB 是随机选取的,要想重构作品的原始版本并不可行。但进行复制攻击仍然很容易:只需将 x_1 的所有 LSB 全部复制到目标作品 x_2 的 LSB 位置上即可。这样一来,虽然 LSB 水印算法能阻止对手重构原始作品,但并不能抵抗复制攻击。

8.4.3 多重嵌入攻击

多重嵌入攻击方法就是对手在别人的含水印图像中加入自己的水印,若做得好这个操作可引起微弱的质量下降。攻击者可方便地用软件实现这个操作以迷惑拥有者。这里需要强调秘密水印方案和公开水印方案的重要区别:公开水印与秘密水印相比,它容易被另一个水印覆盖而删除。在信息伪装系统中,秘密水印的秘密信息可选择隐藏信道,在大量的可能位置中隐藏水印。这就允许嵌入两个水印,每一个按不同的密钥嵌入,它们在不同的位置而不会互相破坏。而公开水印则必须放在检测器知道的每一个地方,使用两个水印就会占用同一个“空间”。

一些公开水印软件,像 Digimarc 公司的水印软件 PictureMarc 就认识到了这个问题,它拒绝让一个用户在水印已经存在的情况下再加水印。但是,可用 Adobe Photoshop 分层功能来进行去除攻击。拿一幅在 Photoshop 中用 Digimarc 的插件程序加入水印的图像,要想直接加入另一个水印已经不允许。但是,可以先对原含水印图像进行连续模糊化直到旧水印提取不出来,然后可以在模糊化后的图像上以最大强度加入新水印(只是简单的覆盖它,以 30%的透明度加在原始图像的上面)。利用 Adobe Photoshop,把背景图像放入一个层内,模糊化的含水印图像放在另外一层,叠加在一起,设置其透明度,并将其平整化。结果看起来好像只是微小的模糊化,这种效果可通过锐化滤波来修复。这样一来,所生成的图像中包括加入的高强度新水印,而旧水印则不再能被检测到。

由前面描述可知,许多秘密水印能够抵抗二次嵌入,因为水印嵌入的方式是保密的,或是水印嵌入的位置是保密的,水印所占的“空间”足够大,以至于攻击者在每个可能隐藏水印的地方都要嵌入自己的水印,一定会对内容造成很大破坏。而对于公开水印,人们会有这样的疑问:如果可加入第二个水印,那么对手就能宣称具有对内容的所有权吗?答案是否定的。这一点之所以不能实现是因为,原始内容的创建者具有真正的原始内容,而攻击者不知道。它不包含水印,而攻击者所谓的原始内容包含第一次加入的水印,而不包含第二次加入的水印,这样自然就有了嵌入先后次序问题。一般情况下,多重水印的怀疑可通过每一方在另一方的正版原图中提取自己的水印来解决。如果每一个原图都是另一个原图的水印版本,那就太奇怪了。如果水印强度非常接近的话,就会阻止任何一方确立所有权。由于多重嵌入攻击存在上述问题,通常对手采用另一种更有效的攻击方法,即协议攻击。在这类攻击中,对手通过伪造原始载体和水印的方法,造成版权解释不清,具体见下一小节。

8.4.4 协议攻击

协议攻击(Protocol Attack),也称解释攻击(Interpretation Attack)可以制造出在作品中嵌有水印的伪象,而实际上并未发生这种嵌入。对手可使用这种攻击来对所分发的作品声明版权,甚至还能够对原始作品声明版权。解释攻击既可看作一种非授权嵌入攻

击，也可看作系统攻击。

1. 明检测解释攻击

明检测解释攻击也就是所谓的 IBM 攻击，即针对可逆、非盲水印算法而进行的攻击。文献[150]中提到的解释攻击对使用明检测器的系统很有效，其原理描述如下：对手 Bob，将其伪造的水印定义为随机生成的参考模式。然后，他将此参考模式从 Alice 分发的含水水印作品中减去，生成他的伪造原始作品。虽然伪造的水印同 Alice 所分发的作品的相关性很小，但它同被分发作品与伪造原始作品之间的差异却有很大的相关性。若设 Bob 所选择的参考模式同 Alice 选择的不相关（这种可能性极大），那么 Alice 的真实原始作品同 Bob 所伪造的原始作品之间的差异也会同伪造水印高度相关。这便使 Alice 和 Bob 都能够对所有权进行声明。虽然 Alice 能够在被分发作品和 Bob 的“原始”作品中检测到她的水印，从而声明二者都是其原始作品的后代。但是，Bob 同样也能这么做，因为 Bob 的水印也能从被分发作品和 Alice 的真实原始作品中检测到。图 8.1 描述了这个过程，Alice 把水印 w_a 加到图像 x 中得 $x^w = x + w_a$ ，然后发布到网上。Bob 想把它据为己有，他从图像 x^w 中减掉（而不是加入）自己的水印 w_f ，得到图像 $x' = x + w_a - w_f$ 。然后，Bob 声称 x' 是他的原始图像，反而把 Alice 拉上法庭告她侵权。通过仔细分析，Alice 会发现她的水印 w_a 仍在 Bob 的图像 x' 中，即

$$x' - x = w_a - w_f \quad c(w_a - w_f, w_a) = 1 \quad (8.24)$$

式中，函数 $c(w, \hat{w})$ 表示 w 和 \hat{w} 是否相似（1 表示相似，0 表示不相似）。既然两个水印可在同一个图像中存在，那减掉 w_f 应该对 w_a 的伤害不是很大。虽然 Bob 并没有成功删除 Alice 的水印，但是 Bob 也可证明

$$x - x' = w_f - w_a \quad c(w_f - w_a, w_f) = 1 \quad (8.25)$$

换句话说，Bob 的水印也在 Alice 的原始图像中，即使 Alice 没有把它公开。

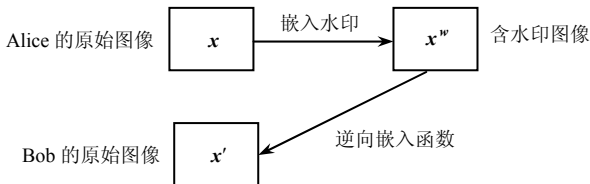


图 8.1 伪造原始图像

2. 盲检测解释攻击

针对盲检测的解释攻击^[5]可通过构造形似噪声但却同所分发作品高度相关的伪造水印来进行。这种参考模式可通过抽取并改变被分发作品的一些特征来建立。这种参考模式可在被分发作品中找到，也极可能在原始作品中找到。对手将此模式从被分发作品中减去，得到他伪造的原始作品，然后再将伪造原始作品和伪造水印锁在保险箱内。文献[5]给出一个实际攻击方法，主要思想如下：创建伪造水印使其同被分发图像之间高度相关，但看上去像噪声，这里对手可用如下步骤产生伪造水印：① 在被分发图像上添加随机噪声。② 计算含噪声图像的傅里叶变换。③ 对傅里叶变换系数进行缩放，使其具有随机幅度（同原始幅度无关）。④ 作傅里叶逆变换并将系数归一化在 1 之内。实验表明，即使这个模式看上去很像纯白噪声，它同被分发图像之间的相关值却能达到 0.968。对手现在必须创建他的伪造原始图像。他可简单地将伪造水印从被分发作品中减去，但这样得到的“原始”图像刚好同伪造水印正交。虽然某个随机模式可能会同原始图像之

间的相关度极低,但这个相关值不太可能正好是零。若参考模式同原始图像之间相关值为零,则对手对所有权的声明便显得十分可疑。为此,对手可只将部分伪造水印从被分发图像中减去,得到伪造的原始图像。实验表明通过从被分发图像中减去 99.5% 的伪造水印,就能攻击成功,也就是说单独使用水印检测器并不能解决真实所有者同对手之间的所有权声明纠纷。

8.4.5 针对脆弱水印的非授权嵌入攻击

1. 生日攻击

文献[151]提出了针对文献[152]的一种图像认证算法的攻击方法。由于文献[152]利用密码学的 Hash 函数,而生日攻击是 Hash 函数的一种重要攻击方法,故文献[151]提出了相应的生日攻击方法。下面,介绍 Hash 函数及其通用的生日攻击思想,限于篇幅不再叙述文献[151]中针对 Wong 算法的攻击过程。

密码学上的 Hash 函数是一种将任意长度的消息压缩到某一固定长度消息摘要的函数。Hash 函数可用于数字签名、消息的完整性检测、消息的来源认证和检测等。安全的 Hash 函数的存在性依赖于单项函数的存在性。也就是说,已知 Hash 函数值,构造一个消息,使其 Hash 函数值相同,应具有计算复杂性意义下的不可行性。

评价 Hash 方案的一个办法是估计找到一对碰撞消息所花的代价有多高。攻击 Hash 函数的主要目标是找到一对或更多对碰撞消息。在目前已有的 Hash 攻击方案中,一些是一般的方法,可攻击任何类型的 Hash 方案,例如生日攻击方法;另一些是特殊的方法,只能用于攻击某些特殊类型的 Hash 方案,例如适用于攻击具有分组链结构的 Hash 方案的中间相遇攻击,适用于攻击基于模算术的 Hash 函数的修正分组攻击。

生日攻击方法没有利用 Hash 函数的结构和任何代数弱性质,它只依赖于消息摘要的长度,即 Hash 值的长度。这种攻击对 Hash 函数提出了一个必要的安全条件,即消息摘要必须足够长。生日攻击这个术语来自于所谓的生日问题,在一个教室中最少应有多少学生才使得至少有两个学生的生日在同一天概率不小于 $1/2$? 这个问题的答案为 23。下面详细描述生日攻击的方法。

设 $h: X \rightarrow Y$ 是一个 Hash 函数, X 和 Y 都是有限的,并且 $|X| \geq 2|Y|$, 记 $|X| = m$, $|Y| = n$ 。显然至少有 n 个碰撞,问题是如何去找到这些碰撞。一个很自然的方法是随机选择 k 个不同的元素 $x_1, x_2, \dots, x_k \in X$, 计算 $y_i = h(x_i)$, $1 \leq i \leq k$, 然后确定是否有一个碰撞发生。这个过程类似于把 k 个球随机地扔到 n 个箱子里边,然后检查是否某一箱子里边至少有两个球。 k 个球对应于 k 个随机数 x_1, x_2, \dots, x_k , n 个箱子对应于 Y 中的 n 个可能的元素。计算用这种方法找到一个碰撞的概率下界,该下界只依赖于 k 和 n , 而不依赖于 m 。因为人们关心的是碰撞概率下界,故可假定对所有 $y \in Y$, 有 $|h^{-1}(y)| \approx m/n$ 。这个假定是合理的,这是因为如果原像集 $h^{-1}(y)$ ($y \in Y$) 不是近似相等的,那么找到一个碰撞的概率将增大。

因为原像集 $h^{-1}(y)$ ($y \in Y$) 的个数都近似相等,并且 x_i ($1 \leq i \leq k$) 是随机选择的,所以可将 $y_i = h(x_i)$, $1 \leq i \leq k$ 视作 Y 中的随机元素 (y_i , $1 \leq i \leq k$ 未必不同)。但计算 k 个随机元素 y_i , $1 \leq i \leq k$ 有不同的概率是一件容易的事情。依次考虑 y_1, y_2, \dots, y_k 。 y_1 可任意地选择; $y_2 \neq y_1$ 的概率为 $1 - \frac{1}{n}$; $y_3 \neq y_1, y_2$ 的概率为 $1 - \frac{2}{n}$; \dots ; $y_k \neq y_1, y_2, \dots, y_{k-1}$ 的概率

为 $1 - \frac{k-1}{n}$ 。因此，没有碰撞的概率是

$$(1 - \frac{1}{n})(1 - \frac{2}{n}) \dots (1 - \frac{k-1}{n}) \quad (8.26)$$

众所周知，若 x 是一个比较小的实数，那么 $1 - x \approx e^{-x}$ 。而上式可估计为

$$(1 - \frac{1}{n})(1 - \frac{2}{n}) \dots (1 - \frac{k-1}{n}) \approx e^{-k(k-1)/2n} \quad (8.27)$$

设 ε 是至少有一个碰撞的概率，则 $\varepsilon \approx 1 - e^{-k(k-1)/2n}$ ，从而有 $k^2 - k \approx n \ln(1/(1-\varepsilon)^2)$ 。去掉 k 这一项，可得 $k^2 \approx n \ln(1/(1-\varepsilon)^2)$ ，即

$$k \approx \sqrt{n \ln(1/(1-\varepsilon)^2)} \quad (8.28)$$

如果取 $\varepsilon = 0.5$ ，那么 $k \approx 1.17\sqrt{n}$ 。这表明，仅 \sqrt{n} 个 X 的随机元素就能以 50% 的概率产生一个碰撞。注意 ε 的不同选择将导致一个不同的常数因子，但 k 与 \sqrt{n} 仍成正比例。如果 X 是一个教室中的所有学生的集合， Y 是一个非闰年的 365 天的集合， $h(x)$ 表示学生 x 的生日，这时 $n=365$ ， $\varepsilon = 0.5$ ，由 $k \approx 1.17\sqrt{n}$ 可知， $k \approx 22.3$ 。因此，此生日问题的答案为 23。

生日攻击隐含着消息摘要长度的一个下界。一个 40 比特长的消息摘要是很不安全的，因为仅仅用 2^{20} （大约一百万）次随机 Hash 可至少以 0.5 的概率找到一个碰撞。为抵抗生日攻击，通常建议消息摘要的长度至少应取为 128 比特，此时生日攻击需要约 2^{64} 次 Hash。安全的 Hash 标准的输出长度选为 160 比特正是出于这种考虑的。

2. 复制攻击

首先需要指出，一般文献中提到的拷贝攻击都是针对鲁棒水印算法。实际上，这种攻击在脆弱水印的某些应用场合中同样适用。文献[153]中提到了针对脆弱水印的复制攻击方法，并针对两种特定的图像认证算法提出相应的攻击方法和实验结果。在基于分块的脆弱水印应用场合，若已知在图像 x^w 中包含水印 w ，则复制攻击过程可统一描述如下：输入为含水印 w 的图像 $x^w = \{x_1^w, x_2^w, \dots, x_n^w\}$ 和不含水印的图像 $y = \{y_1, y_2, \dots, y_n\}$ ；输出为包含水印 w 的图像 $y^w = \{y_1^w, y_2^w, \dots, y_n^w\}$ ，使得 $y^w \cong y$ 。对于第 k 块，操作过程为：

① 识别块 x_k^w 所属的类 C_k ；② 构造接近 y_k 的块 y_k^w 使得 $y_k^w \in C_k$ ；③ 用 y_k^w 替换 y_k 。

显然，如果攻击者知道水印，他至少也知道对应给定图像块的等价类 C_k 的部分知识。这样一来，可利用图像矢量量化技术来构造 y_k 的逼近块 y_k^w ，使其属于等价类 C_k ，对应的矢量量化码书包含所有属于类 C_k 的 x^w 中的含水印块。若攻击者不知道水印，则他可利用水印信号结构的部分知识来伪造水印。限于篇幅，在此不再赘述。

3. 二次拍摄攻击

除了上面两种方法，文献[138]还提出一种二次拍摄攻击。其基本思想如下：对含水印载体进行篡改后进行二次拍摄，在拍摄过程中以前的脆弱水印丢失，在拍摄时加入新的脆弱水印，这样经过篡改的图像仍然认为是真的。

8.4.6 对策

下面考虑如何对付非授权嵌入的问题，对前三类攻击的对策分别加以说明。

1. 针对协议攻击的对策

首先需要强调：要阻止对手使用这种协议攻击来创建伪造的原始作品和伪造的水印是不可能的。防止这类攻击的主要手段依赖于：作品的真实所有者必需使用一种水印技术，使其能够拥有比对手更有力的证据来支持其声明。实际上，协议攻击所利用的安全漏洞是水印算法的可逆性（Invertibility）^[150]。如果嵌入过程的逆过程在计算上容易实现，则把水印算法称作是可逆的（Invertible）（需要指出，这个可逆性是针对计算复杂度而言的，而不是第5章中的含义）。嵌入过程的逆过程是被分发作品 \mathbf{x}^w 的函数，能够生成一件伪造的作品 \mathbf{x}' 和伪造的水印 \mathbf{w}_f ，使得嵌入函数 $\varepsilon(\cdot)$ 能够将伪造的水印嵌入到伪造的原始作品中，得到当前的被分发作品。用符号描述这个过程，设嵌入函数为 $\varepsilon(\mathbf{w}, \mathbf{x})$ ，存在如下等式

$$\varepsilon^{-1}(\mathbf{w}_f, \mathbf{x}^w) = \mathbf{x}' \quad (8.29)$$

使得

$$\varepsilon(\mathbf{w}_f, \mathbf{x}') = \mathbf{x}^w \quad (8.30)$$

解释攻击不会对不可逆（Non-invertible）的嵌入技术（即嵌入过程的逆过程在计算上不可实现）起作用。创建这种不可逆嵌入器的一个方法是使参考模式依赖于原始作品的内容，以使这些参考模式无法在缺少原始作品的情况下生成。利用这种依赖性，可使用单向 Hash 函数来保证对手无法伪造一个能够生成正确伪造水印的原始作品。例如，可通过原始图像 Hash 值作种子的伪随机噪声发生器来产生一个水印参考模式。对手可在有限时间内找到一个同被分发作品高度相关的随机水印，但这个随机水印却不是以伪造的原始图像的 Hash 值作种子的伪随机噪声发生器所生成的。也就是说，想让通过伪造的原始作品的 Hash 值伪随机生成的水印同被分发作品高度相关几乎是不可能的。

2. 针对复制攻击的对策

然而，上述非可逆水印方法并不能解决第二类非授权嵌入攻击，对手从合法含有水印的作品中提取出水印并将其复制到一件不含水印的目标作品中。如果水印已被一私钥加密，则其加密后的形式可被全部复制，因而也能够用对应的公钥解码。如果水印具有加密签名，能够随同水印被复制，那么签名仍然会同消息的 Hash 函数相吻合。这种作假行为看上去完全合法。原因在于，嵌入作品的消息确实是真实的，但其隐藏作品却不对。根据定义，水印必须携带所嵌入作品的信息，同作品无关的消息很难被正确理解。在图像中嵌入“你可以复制”的消息意味着你被授权对此图像而不是其他图像进行复制。严格地讲，含水印作品必须看成消息的一个隐含部分。为抵抗这种攻击，Alice 必须能够对整个水印消息进行验证，包括其同含水印作品之间的关联性。

怎样验证这种关联性呢？一种可能是在要嵌入的消息上附加隐藏作品的一个描述。一种显然的方法是：在计算加密签名之前，将整个含水印作品附加在消息上。然后将消息随同签名一齐嵌入。然而，此方法并不可行，因为嵌入过程对作品进行了改动，会使签名失效。为避免这个问题，可对作品的某一部分进行标记，例如作品的低频成分，而不是整幅作品。如果嵌入过程不改动作品的这一部分，那么签名在水印嵌入后依然有效。这样，为嵌入一个水印，Bob 必须进行下列操作步骤：根据含水印作品不太可能变动的信息，如低频分量，生成作品的一个描述。将水印消息连同含水印作品的描述一齐生成一个单向 Hash 值。对此 Hash 值用私钥进行加密，得到加密签名。使用不会改变前面描述的嵌入算法，对水印消息随同签名进行嵌入。为检测和验证水印，Alice 须进行下

列操作：检测水印并对其解码，得到消息和加密签名；生成同 Bob 相同的含水印作品的描述；计算水印消息连同含水印作品描述的单向 Hash；使用 Bob 的公钥对加密签名进行解码；比较解码的签名和消息及含水印作品描述的 Hash 值，如果两者相同，Alice 便能确认 Bob 生成了此消息并将其嵌入在此作品当中。

这种系统存在一个难题：如果作品在嵌入操作和检测操作之间发生质量恶化，那么水印签名用到的描述必须极其鲁棒。然而，即使发生 1 比特的变化，单向 Hash 值将会完全改变，签名会不再有效。为解决此问题，可以使用一种不同的方法。Bob 仍然计算这件作品的描述，并将其用在加密签名中。但在嵌入消息和签名的同时，Bob 还将嵌入描述本身。之后 Alice 可得到 Bob 用来计算签名的那份作品描述。她使用这个签名验证所有嵌入的信息（即消息和描述）全部有效，接着对所嵌入的描述和接收作品产生的描述进行非精确比较。例如，她可以计算这两个描述之间的相关，并将结果同阈值相比较。这个系统允许 Alice 算得的描述与 Bob 算得的描述略有不同，而不会认为签名无效。

无论在哪一个系统中，作品的描述应能够反映出作品绝大部分重要感知特征。准确地说，必须保证极难从两件感知上不同的作品得到相同的描述。否则，对手在将合法作品中的水印复制到目标作品中时，可通过修改目标作品以使其描述同合法作品的描述相匹配，从而使签名失效。

上述联系水印同其隐藏作品的描述系统反映出网络层次模拟结构的局限。这种通信层次模型的特点在于层次之间相互独立，只需在层与层之间清晰地定义好接口。然而，对于水印处理而言，这类系统意味着如果人们想在作品和消息之间建立联系，消息（加密）层必须要根据所使用的特定的传输（水印）层来设置。首先，它必须将隐藏作品的知识融入消息的加密保护过程中；另外，它必须根据具体的水印算法来确定作品的一些性质，以保证嵌入过程不会对这些性质产生影响。

3. 针对多重嵌入攻击的对策

前面已经提到，多重嵌入攻击并不能通过加入第二个水印宣称具有对内容的所有权。这一点之所以不能实现是因为，原始内容的创建者具有真正的原始内容，而攻击者不知道。它不包含水印，而攻击者所谓的原始内容包含第一次加入的水印，而不包含第二次加入的水印，这样自然就有了嵌入先后次序问题。但为了说明对策，这里假设这种攻击是有威胁的。这种攻击可用以下两种方法解决：

(1) 最大强度嵌入。也就是说，原始内容的创建者在嵌入水印时要在保证不可见性的同时嵌入最大能量的水印，以使第二次嵌入一定会影响图像质量。

(2) 时间戳 (Time Stamps)。为了确定谁第一个给图像作了标记，就需要用到时间戳（由可信赖的第三方提供）。令 \mathbf{x} 为将要加入时间戳的图像， H 为对应的散列值。

所有者发送一个正式请求 $R_n = (H_n, s_n)$ 到一个正式的第三方时间戳服务机构 (TSS)，此处 s_n 是所有者的标识字符串。TSS 产生一个时间戳 TS_n 如下

$$TS_n = f_k(n, s_n, H_n, T_n, H_{n-1}, T_{n-1}, L_n) \quad (8.31)$$

此处 n 是请求数目， T_n 是请求时间， f_k 表示信息已用 TSS 的公钥进行了签名。 L 被称为连接串，由下式定义

$$L_n = H(s_{n-1}, H_{n-1}, T_{n-1}, L_{n-1}) \quad (8.32)$$

上式用于避免因时间戳申请者和 TSS 串通而产生他们需要的任何时间戳。TSS 然后等待下一个请求并且回复新的原始请求者标识 s_{n+1} 。如果有人质询一个时间戳 TS_n ，则所有者可证明时间戳是在 s_{n-1} 之后与 s_{n+1} 之前盖的。如果他们给出的文档也受到怀疑，则

他们可使用 s_{n-2} 和 s_{n-1} ，以此类推。由于数字时间戳涉及到一个可信赖的第三方，这就会产生一个为什么将水印和时间戳结合起来使用的疑问，因为这和传统的版权注册和版权保护法律非常相似。

8.5 非授权检测攻击

8.5.1 问题

在一些检测操作应受限制的应用场合，人们主要关心怎样防止对手对水印消息的解码。例如，医院可将病人的姓名嵌入到 X 光片中。为保护个人隐私，非授权的个人应禁止读懂这些印记和识别病人身份。

而在某些应用场合，对手可能只满足于得知水印是否存在，而并不关心如何对水印进行解码得到最终的消息。严格说来，这是一个与隐写术而非水印相关的问题。但仍然可以设想一个水印应用场合，对手可仅仅通过判断作品中是否存在水印而破坏水印系统的安全性。另外，能够检测出水印的存在对于想要去除这些印记的对手而言是一个很大的帮助。因此，可以认为：若要保证水印具有对非授权去除攻击的安全性，就必须保证其具有对非授权检测的安全性。仅仅具有防止对水印进行解码的安全性是不够的，还必须还具备针对检测攻击的安全性。

非授权检测的第三种类型介于全解码和存在性检测之间，此时对手可分辨出用于对不同消息编码的所有印记，即使他可能并不知道这些印记所编码的具体消息是什么。也就是说，如果拿给对手两件嵌有水印的作品，他可以确定说出两个水印是否是相同消息的编码形式。如果对手可以通过解码以外的其他方式判断出消息含义的话，这种类型的攻击确实是一个很大的隐患。例如，在 8.2.2 小节的场景 2 下，Bob 希望能够具有找出水印中所嵌 ID 信息的能力，这样他就可从 Alice 那儿抢走她的客户。Alice 可能会试图通过加密编码的方式将水印同客户信息关联起来以防止 Bob 通过解码而盗取作品的所有者信息。然而，如果 Bob 能够可靠地确定是否两件作品都含有同一水印的话，他仍然能抢走 Alice 的客户。Bob 只需要向任何他的未来客户索取一件经 Alice 加入水印的作品就能办到这一点。当他再在网上发现含有相同水印的其他作品时，他便得知其属于此客户所有。这种情况下，Bob 并不需要知道 Alice 怎样把每一客户的信息编码成水印的。

8.5.2 对策

下面介绍如何防止非授权检测。首先介绍 Alice 和 Bob 之间秘密通信的问题。在水印处理中，要求他们防止对水印消息的非授权检测和解码。理想情况下，他们总希望使用对这些攻击类别具有安全性的水印处理系统。然而要设计一个完全安全、能抵抗非授权检测和解码攻击的水印处理系统是不太实际的。一般而言，如果水印系统要具备对这些攻击的安全性，密钥空间必须非常大。就是说，必须存在一个很大的密钥集合使得 Alice 和 Bob 可从中选取。如果密钥空间过小，对手（Eve）便能够通过强力（Brute-Force）搜索找出正确的密钥（假定她知道水印算法，这是最安全的假设）。困难在于水印算法的设计往往需要在许多矛盾的要求下取得折衷，这可能导致密钥空间很小。在这种情况下，非授权解码的问题可通过直接应用加密算法来解决。可以对系统使用传统的加密算法。消息在嵌入前进行加密，检测后进行解密。这种系统需要两个密钥：水印嵌

入密钥 K_w ，控制水印嵌入这一层，以及水印生成密钥 K_c ，控制消息加密层。在嵌入端，含水印作品可由下式给出

$$\mathbf{x}^w = \varepsilon_{K_w}(\mathbf{x}, \mathbf{m}_c) = \varepsilon_{K_w}(\mathbf{x}, E_{K_c}(\mathbf{m})) \quad (8.33)$$

在检测端，应用相反的过程

$$\hat{\mathbf{m}} = D_{K_c}(D_{K_w}(\hat{\mathbf{x}})) \quad (8.34)$$

当水印使用符号序列对消息进行编码时，系统的加密层与水印层的功能划分可以这样描述：加密层隐藏消息，而水印层隐藏符号。加密层和水印层可看成网络系统的两层。网络系统一般被划分为数层，每一层负责一项具体的工作。这里所述的两层分别为消息层（Message Layer），用于确定在网络上传输的消息，以及传输层（Transport Layer），保证所传输的消息能安全到达。加密是消息层的一部分，水印系统连同隐藏作品则是传输层的一部分。

增加一个加密层可保证消息的安全性，即被检测出来的水印的含义无法被破解。另外，在一些系统中可防止对手确定水印的存在性。如果所设计的水印系统通过区分消息的真实性来确定水印的存在性，那么不通过解密就不可能识别出真实的消息。这样，拥有嵌入密钥而不具有生成密钥的对手就无法确定水印的存在性。然而在绝大多数水印系统中，加密并不能防止被编码消息被检测出来。在一些系统中，可通过对检测统计量和相应阈值进行比较来判断消息的存在性。在另一些系统中，水印的存在性也可通过其他一些方法进行检测。例如，在作品的最低有效位嵌入水印会在作品的直方图上产生指示性的统计失真效应。这些失真可被用来区分含水印作品和不含水印作品，即使这些印记相对噪声而言并不明显。

从消息层 / 传输层模型中可确切地看出加密方法的局限。在未解密的情况下如何防止非授权检测是传输安全的问题，因为对手原则上一定希望检测出消息正在传输这个事实。而加密是消息层的一部分，同消息的保密性、真实性、完整性和不可否认性相关，因此非授权检测不太可能通过加密技术来抵御。

8.6 系统攻击和法律攻击

8.6.1 引言

在结束攻击的讨论之前，必须指出并非所有的攻击都是针对水印本身的。对手经常能够对系统进行破坏而并不需要利用前述的几种非授权行为。把那些利用水印使用上的弱点而不是水印本身弱点的攻击行为统称为**系统攻击**（System Attack）。系统攻击也称**合法攻击**（Legal Attack）或法律攻击，因为这种攻击主要是利用法律上的一些条款的漏洞以达到攻击的目的，以破坏作为所有权证据的水印的可信性为目的。换句话说，这种攻击不包括对水印作品的伪造，而是试图利用水印作为所有权证据的法律基础上的缺陷（如版权法立法上的缺陷），对所有者的可信性进行挑战。这种攻击已经超出技术范畴，本节只做简单介绍。

考虑系统攻击的一个简单的例子，在复制控制应用中，每个录制设备中都安装有用于检测水印的计算机芯片。对手可能只需打开这个录制设备，取出芯片，即能使此录制设备不再禁止非法复制。这种攻击与水印本身的安全性根本无关。当然，本节仅讨论水印技术，因而诸如拔除检测器芯片之类的系统攻击不在讨论范围之内。然而，在设计水

印应用系统时，考虑此类系统攻击则非常重要。对手总是会攻击安全链条中最薄弱的环节。本节所考虑的攻击都基于信号处理，首先讨论由系统设计失误引起的水印算法缺陷，如系统设计没有考虑人为因素、用户接口和实现上的缺陷等，然后介绍典型的法律攻击。

8.6.2 体系结构问题

1. 人为因素

典型的用户通常对图像水印只是有限地了解，不想花很多的时间去使用图像处理软件中的某个函数。水印应该像“黑箱子”，而用户输入原始图像，通过某种“魔力箱子”输出含水印图像，用户不必要理解具体的细节。

通常，画家和设计者创建的图像需要版权保护。画家不想因为加入水印后降低画的质量，故画家要么加入弱水印要么不加任何东西。很难说服画家们仅仅为了安全性而降低他们作品的图像质量。

2. 用户接口

考虑到上面所描述的人为因素，可以看到**用户接口**（User Interface, UI）成为安全体系结构的重要组成部分。UI 应该为用户提供一个水印效果的清楚模型，防止用户误用。这需要消除用户对水印技术的错误概念。

因为水印嵌入程序通常是与图像处理应用在一起的，所以用户在处理图像时可能会误删水印。在当前的应用中，比如说 Adobe Photoshop 的 PictureMarc，这个问题就没有解决，用户不清楚水印被减弱是因为改变图像颜色、修改边界或者是什么其他操作导致的。很明显，UI 应该防止用户偶然删掉或削弱水印。一种可能的改进就是在屏幕上放置一个水印强度的标示器，这样用户可以看到水印对图像变换是怎样反应的。理论上水印嵌入应该是图像出版之前的最后一步，因此软件可以将嵌入过程推迟直到用户要保存图像之前。同样在下载图像时，水印可以先提取出来，用户再对不含水印的图像进行操作。这样，水印将变成一个对用户透明的操作。

3. 实现过程的缺陷

大多数对密码系统的攻击来自于攻击者对偶然发现的漏洞的利用，即使在攻击很脆弱的系统时，也很少使用密码分析学。人们不能期望版权水印系统有什么不同，在因特网上对最广泛应用的图像水印方案的攻击中，这种方法也会被使用。该攻击方法利用了实现中的缺陷，而不是水印算法，即使算法是脆弱的。

在互联网上最广泛应用的图像水印方案中，每个用户有一个 ID 和一个两位口令，这是用户在向水印软件供应商注册时所得到的。在嵌入水印时，检查 ID 和口令是否相符。人们以为这种检查可以防止攻击者用已知 ID 嵌入水印，但不幸的是，设计上的缺陷允许攻击者用两种不同的方法攻破系统。

第一种攻破系统的方法是：用调试程序攻破软件，使口令检查机制不能正常工作。这种攻击可以从网上得到。第二种方法是与 ID 相匹配的秘密口令只有两个十进制位长，所以只有一百种可能，平均花一百秒就可以找到任何用户 ID 的秘密口令！更别提这个过程自动完成多容易了。顺便提一下，如果 ID 公开，那么口令的搜索或分解将使任何一个用户都能被冒充。

通过对这个程序更深入地分析，攻击者还可以修改 ID、含水印图像的版权以及应用类型。在嵌入水印之前，程序先检查图片中是否有水印，但这个检查可以通过调试程序

很轻松地跳过，结果攻击者可能用另一个水印覆盖任何已存在的水印。

对个人口令的穷举搜索可以通过增加搜索空间来防止，但对分解攻击则没有有效的解决方法。如果防篡改软件不能给予足够的保护，那么人们可以用在线系统，让每个用户和信任方共享一个保密的嵌入密钥，用这个密钥嵌入某种数字签名。这样，在水印系统中将有两种相互分离的带密钥的操作，即身份认证（可以用签名来实现）和嵌入或隐藏操作。

4. 自动蜘蛛限制

在前面我们讨论了马赛克攻击（虽然我们将它归入非授权去除攻击中的掩盖攻击，实际上它也可看作系统攻击的一种），它可以阻止 Web 蜘蛛检测到图像中的水印，即使图像外表并没改变。但不幸的是，Web 蜘蛛在网上搜索盗版图像时还有更多问题存在。

第一个问题是带宽：在因特网上爬行需要很高的带宽。这使得个人用户在网上搜索自己的被侵权的图像是不可能的，因此用户需要在公共服务上注册来寻找图像。如果他使用秘密水印，他将需要用蜘蛛注册私钥，这会引起很多密钥管理和安全问题。

现在让我们看看整个周期，分析一下会遇到的问题。设想 Alice 是一个画家，她创造了持有版权的有价值的图像。嵌入秘密水印后，她在 Mallory 上注册检测侵权，并把密钥给了他。在这个情景中，Mallory 已经偷了 Alice 的图像并自豪地放到了自己的网页上。因为 Mallory 知道 Alice 拥有这些图像的版权，所以他编程让他的 Web 服务器对 Web 蜘蛛不提供针对所偷图像的服务。假设没有很多公司上网，这个防御措施就很容易实现。为避免猜疑，当 Web 蜘蛛对被偷图像提出要求时，服务器可返回一个任意图像。

Web 蜘蛛的另一个缺陷是对于访问控制网站或付费站点，蜘蛛没有身份认证或信息付费就不能访问图像。我们不知道哪个法律能允许蜘蛛或政府代理为了调查而免费搜索访问控制站点。即使有这样的法律，网络服务器也可以检测这样的用户，给他一个假冒的图像。不幸的是，访问控制（付费）站点上的侵权大多直接伤害内容的创建者，因为他们是从创建者作品中挣钱的。如果蜘蛛在检查图像之前不得不用信用卡买下图像，那他就成了网站的目标，提供“有保证的销售”了。这些问题都是摆在 Web 蜘蛛面前要克服的障碍。

Java 应用程序或 ActiveX 控件控制那些能嵌入浏览器显示图片的动态对象，因此他们为蜘蛛又提出了一个大问题，Java 程序甚至可以实时地解码图片。打败这种技术将需要修补网页、检测图片和检测它们是否有水印。

另外，Mallory 甚至可以不真正偷图像而将 Alice 的图像放到自己的网页上。通过 HTML，Mallory 可以不用将图像复制到自己的网上空间，而是在 Alice 的服务器上直接访问图像。然后参观者的浏览器从 Mallory 的服务器上下载网页，从 Alice 的服务器上下载图像。人们相信 Alice 控告侵权将很困难，因为图像只有在她的服务器上才找得到。

8.6.3 典型合法攻击

上面介绍了可以阻止 Web 蜘蛛找到被盗图像的很多方法。下面考虑侵权被检测到之后一些问题。

第一种情况是：在一个没有加入有关版权保护的伯尔尼协定的国家，Mallory 建立自己的网络服务器发布侵权的图像、音乐等。没有办法阻止他做这些非法发布，因为该国不提供起诉的合法法律基础。所有的知识产权形式都有这个问题，尤其是当互联网迅速发展，该情况也日益严重。过去的事实说明这种情况并不是牵强附会的。事实上，最近的很多事件证明了这种攻击是相当普遍的。随着因特网的延伸，高速连接到对侵权还处于传统意识上的国家，情况变得更糟糕。如果这些国家还不改变他们的法律的话，这个

问题不能从技术上解决。

第二种情况是：即使在实施版权保护公约的国家里，对版权侵犯实施法律制裁也不是一件简单的事。人们很难在法庭上证明某个版权确实遭到了侵犯，它的难点在于 Alice 并不能因为声称“Mallory 在他的网络服务器上发布了 Alice 版权所有的图像”就认定 Mallory 侵犯了 Alice 的版权。人们需要证据来证明 Mallory 确实有欺骗性行为，通常还要有一个可提供证明的公正的证人。但问题是网络服务器并不能提供不可否认的证据，因此证人无法确认数据确实来自于 Mallory 的服务器。另一种可能的情况是 Mallory 可以试图令 Alice 相信是 Bob 偷了她的图像。而事实上，Bob 是一个好人，不会去偷任何图像。Mallory 可采用的欺骗 Alice 的一种简单方法是进行域名服务器（DNS）欺骗：当 Alice 访问 Mallory 的网页 <http://www.bob-com/image-gif> 时，Mallory 的 DNS 的检查程序用一个错误的地址 <http://www.bob-com> 做出回应，并向 Alice 的浏览器传送那个“被盗用的”图像。还有一种情况是 Mallory 确实盗用了 Alice 的图像，Alice 向法庭提起了诉讼。但在这一过程中，Mallory 完全可以将盗用的图像从他的服务器上删除，并且当 Alice 请公证人检查 Mallory 的网站以证实那里确实存在被盗用的图像时，Mallory 可以拒绝让公证人检查（对某些域名拒绝数据传输在技术上是可行的）。上述这些攻击说明法律系统需要扩展以防止这些问题。

8.7 本章小结

这一章主要讨论了水印的安全性问题，介绍了四大类攻击方法，即非授权嵌入、非授权检测、非授权去除和系统攻击。这些攻击方法指出了现有水印方案中大量的不足之处。由此可见，对于一个水印系统，核心问题是如何检测水印而不是如何嵌入水印。同时应该注意：攻击形式多种多样，并不局限于对水印算法本身，常常一些简单的攻击如马赛克攻击就可以使水印系统失效。了解这些攻击以及可能还会出现的新攻击方法将帮助人们设计出更好的水印方案。为了保证水印消息各个方面的完整性，许多安全性要求可结合加密技术来满足，诸如对消息进行加密嵌入以防止非授权解码。使用非对称加密和加密签名验证水印的嵌入者，防止某种形式的非授权嵌入。使用加密签名验证水印是否属于其所在的隐藏作品。私有水印和公有水印的称谓并不具有私钥和公钥加密的关系，且私有水印并不意味着使用明检测。许多安全要求并不能够直接由加密工具满足。一般说来，下列问题尚无定论仍需研究：要防止未被解码情况下的非授权检测，水印必须不能改变作品的统计特性。要防止掩盖攻击，水印检测器必须能够检测出掩盖失真并将其复原。要防止去除攻击，必须防止对手辨认出检测区域的边界。当对手具有检测器时，这是个尤为困难的问题。



习题

1. 请阐述数字水印系统的鲁棒性和安全性的不同含义。
2. 请阐述数字水印攻击技术的分类。

3. 试用 Matlab 或 C 语言编写一段程序，选择一种空域水印算法在 256 灰度 Lena 图像中嵌入一个二值图标水印，然后对含水印图像进行 4×4 分块，保持边界的含水印图像块不变，随机选择一定数量的中间块，对选中的每一块用其周围的 8 块的平均值替换，

再提取水印，观察水印的破坏程度。请问该攻击方式属于什么攻击。

4. 请解释马赛克攻击的原理。

5. 非授权嵌入攻击方法可以分为哪几类，其基本思想都是什么？

6. 请用 Matlab 或 C 语言编写一段程序，选择一种明检测空域水印算法在 256 灰度 Lena 图像中嵌入一个二值图标水印，来模拟 IBM 攻击的效果：Alice 把水印 w_a 加到图像 x 中得 $x^w = x + w_a$ ，然后发布到网上。Bob 想把它据为己有，他从图像 x^w 中减掉（而不是加入）自己的水印 w_f ，得到图像 $x' = x + w_a - w_f$ 。然后，Bob 声称 x' 是他的原始图像，反而把 Alice 拉上法庭告她侵权。

7. 请阐述针对多重嵌入攻击的对策。

8. 请阐述针对非授权检测攻击的对策。

9. 请举例说明数字水印系统可能受到的系统攻击。

信息隐藏技术的应用

本章引言

基于前面八章内容的介绍，信息隐藏技术在信息安全相关领域的诸多方面发挥着重要作用。总的来看，信息隐藏的应用主要可归结为下列几个方面：① 数据保密通信。通信双方将秘密信息隐藏在数字载体中，通过公开信道进行传递。② 身份认证。利用信息隐藏技术将各自的身份标记隐藏到要发送的载体中，以此确认其身份。③ 数字作品的版权保护与盗版追踪。服务提供商在向用户发放作品的同时，将服务商和用户的识别信息以水印的形式隐藏在作品中，这种水印从理论上讲是不能被移除的。当发现数字作品在非法传播时，可以通过提取的识别信息追查非法传播者。④ 完整性、真实性鉴定与内容恢复。可在数字作品中嵌入基于作品全部信息的恢复水印和基于作品内容的认证水印，由认证水印实施对数字作品完整性和真实性的鉴别并进行篡改区域定位，由恢复水印对所篡改区域实施恢复。本章首先介绍信息隐藏技术的四个最重要的应用领域：知识产权保护（鲁棒数字水印）、军事保密通信（隐写术）、交易跟踪（数字指纹）和真伪鉴别（脆弱数字水印）。然后介绍复制控制、广播监控、设备控制及其他一些应用领域。

本章重点

- 信息隐藏技术在知识产权保护中的应用；
- 信息隐藏技术在军事保密通信中的应用；
- 信息隐藏技术在交易跟踪中的应用；
- 信息隐藏技术在内容认证中的应用。



9.1 知识产权保护

知识产权是指人类智力劳动产生的智力劳动成果所有权。它是依照各国法律赋予符合条件的作者、发明者或成果拥有者在一定期限内享有的独占权利，一般认为它包括版权和工业产权。版权是指著作权人对其文学作品享有的署名、发表、使用以及许可他人使用和获得报酬等的权利；工业产权则是包括发明专利、实用新型专利、外观设计专利、商标、服务标记、厂商名称、货源名称或原产地名称等的独占权利。

数字技术的发展和运用极大地丰富了知识产权客体的范围和种类，数字作品的出现更是彻底颠覆了传统客体以纸制品为主要载体的知识产权时代。数字作品对知识产权的诸多传统特征的巨大冲击，使得数字作品的知识产权保护问题也逐渐突显出来。而且 Internet 发展迅速，尤其是多媒体存储与传输技术的进步，带来了数字媒体应用的迅速增长。多媒体数据的数字化为多媒体信息的存取提供了极大的方便，同时也极大地提高了信息表达的效率和准确性，如数字信号很容易进行编辑，可以方便、便宜、无失真地被复制，数字声音、文本、图像和视频易于通过电子的（网络）或物理的（CD-ROM）系统低价高效地迅速传输和分配等。但是随之而出现的问题也十分严重：作品侵权更加容易。因此怎样更好地保护图像、文本、声音和视频等数字媒体的知识产权就变得十分迫切。一般来说，我们通常会采用三种手段来保护知识产权^[154]：一是法律措施，即通过加强立法，完善有关知识产权保护的法律法规，使知识产权保护有法可依，同时加大司法处理力度，制裁侵权行为；二是行政措施，即通过行政机关或法律授权的裁判组织来处理知识产权纠纷；三是技术措施，也就是说权利人主动采取技术手段，利用技术的保护，阻挡他人的侵权行为。

目前，在数字作品知识产权保护中常用技术主要包括认证技术、加密技术、数字水印、数字指纹、数字签名、数字权利管理、防火墙技术七种。这一节主要考虑基于数字水印的知识产权保护措施。根据应用的不同，基于数字水印的版权保护系统有两种设计方式。一种是数字水印的提取和检测要求实时自动进行，以及时发现数字多媒体数据的非法拷贝；另一种是其嵌入和检测不需要实时自动进行。这与 PKI（Public Key Infrastructure）有共同点（PKI 依靠证书，而数字水印系统依靠水印，两者都需要进行证书或水印的发布和检测），因此人们开始考虑借用公钥基础设施 PKI 的理论和办法，尤其是认证机构 CA 的建设经验，研究数字水印基础设施的建设途径。下面介绍一种基于数字水印的版权保护应用方案^[155]。

9.1.1 基于数字水印的版权保护系统框架

PKI 与数字水印系统的相似之处在于证书管理和水印管理都需要一个权威可信赖公正的第三方机构；不同之处是 PKI 是对证书（密钥）进行管理，数字水印系统是对数字水印进行管理，而且证书和数字水印的用途也不尽相同。数字水印版权系统是一个用数字水印算法和技术来实现并提供多媒体电子商品（作品）安全服务的具有通用性的安全基础设施，扮演着基于数字水印检测、验证和追踪非法复制的角色。下面对该系统框架作详细说明。

1. 角色定义

一个典型的数字水印版权系统包含的角色如下。

(1) 水印管理机构

水印管理机构是其核心组成部分,是权威的、可信的、公正的第三方机构。

(2) 注册机构

确定数字水印和版权所有人之间的连接。

(3) 数字水印信息库

是数字水印信息的集中存放地,提供公众查询。

(4) 水印作废处理系统

水印由于某种原因需要作废或终止使用,通过该系统完成。

(5) 水印撤销列表

记录水印与版权所有人间的断开关系,是作废水印信息的集中存放地,提供公众查询。

(6) 数字水印代理系统

根据应用的不同要求可以有两种形式。一种是“数字水印检测仪”,一种是利用 Agent 技术的网上实时代理。

(7) 应用接口系统

为各种各样的应用提供安全、一致、可信任的方式与其交互。

2. 基于数字水印的版权系统的功能

基于数字水印的版权保护系统应该具有以下功能操作。

(1) 水印的使用

水印的内容一般是由用户自己定义的,系统负责将用户产生的经过水印管理机构签名的水印嵌入到需要保护的多媒体中,或经过审核之后由用户自己嵌入。水印使用的另一个主要内容是水印的检测、提取,这个过程可根据应用的需要实时或非实时进行。

(2) 验证水印

水印 W 在嵌入之前,需要用水印管理机构的根私钥 RootPrivateKey 进行签名,形成消息 $(W, \text{Sign}(W))$,将整个消息嵌入数据中,以证明水印的合法性。当要用水印来确定版权时,可将消息 $(W, \text{Sign}(W))$ 提取出来,并用水印管理机构的公钥 RootPublicKey 验证签名,若验证成功且水印没有作废,则说明水印是合法的,从而证明了版权的合法性。

(3) 保存水印

保存水印是指实体在本地存储水印,以减少在 DWI(数字水印基础设施)体系中对水印的时间。水印存储单元应对存储的水印进行定时管理维护,与最新的水印废止列表相对照,清除已作废的或过期的水印。

(4) 与水印有关的密钥的管理

为了实现系统中各个实体通信的保密性、完整性和相互之间的身份认证,可以利用密码学中的密码技术和算法,这就牵扯到密钥的管理,包括密钥的产生、分发、使用、备份、恢复、废止以及密钥历史记录等内容。

(5) 水印废止的申请

当某用户对某多媒体数据的版权被剥夺或过期时,相应的水印应该被废止。

3. 数字水印基础设施的安全服务

作为安全基础设施,能为不同用户实体提供多种安全服务,分为核心服务和附加服务。

(1) 核心服务

提供的核心服务应该有两个:版权确认和多媒体数据完整性。版权确认是指通过检测多媒体数据中的水印向一个实体证明版权的归属。多媒体数据的完整性是指通过脆弱

水印向一个实体确保数据没有被有意或无意的修改。

(2) 附加服务

附加服务并不是所有数字水印版权系统必须具备的功能。服务包括：复制控制、源指针等。复制控制是指通过嵌入多媒体数据中的水印来限制一个实体对多媒体数据进行拷贝的数量。源指针是指当出现非法数据时，能通过水印确定非法数据的来源。

4. 数字水印管理机构

数字水印管理机构如图 9.1 所示，它包括如下功能模块。

(1) 跟踪、控制、管理中心

实时、动态地跟踪数字水印系统中的其他模块，根据各个模块的执行状态，控制各模块的工作，协调各模块的冲突。

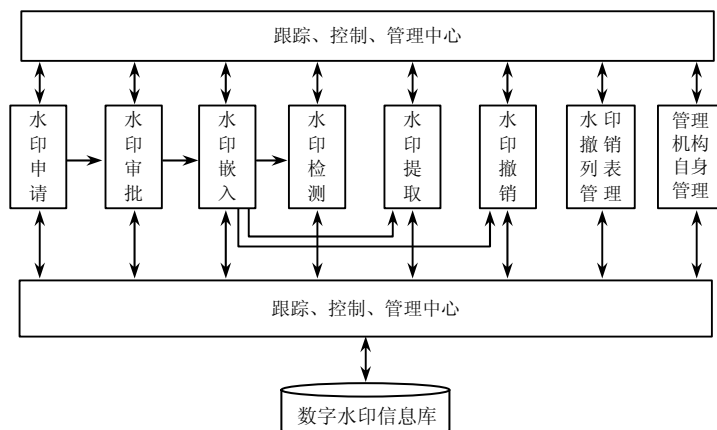


图 9.1 数字水印管理机构

(2) 水印申请

接受申请人的嵌入水印申请。

(3) 水印审批

对申请人的身份进行验证，以确定其是否具备拥有水印的资格。

(4) 水印嵌入

数字水印管理者给每个水印一个唯一的标识（水印序列号），并产生水印嵌入密钥，并将此密钥和水印嵌入程序发给申请人。申请人利用此密钥和程序在数字水印管理者的控制下，将水印信息嵌入原始多媒体数据中。

(5) 水印检测和提取

在生成水印嵌入密钥的同时，数字水印管理者还要生成一个检测 / 提取密钥，根据应用领域的不同，将该密钥只交给申请人或公开给所有人员。得到密钥的人（实体），可以利用相应的检测 / 提取程序对水印进行检测或提取。

(6) 水印撤销

现实世界中的版权是有一定时间限制的，因此作为数字作品版权保护有效手段的数字水印也应该允许撤销。水印的撤销并不是将所有备份中的水印清除，而是断开水印与版权所有人之间的连接。不再主动进行该数字作品的水印检测提取工作。

(7) 水印撤销列表管理

水印撤销列表中记录水印与版权所有人之间的断开关系。

(8) 管理机构自身管理

包括上下级管理机构的关系和不同管理机构之间的互通等问题。

9.1.2 数字水印版权系统与 PKI / CA 体系的对比

由以上分析可以看出数字水印基础设施是借鉴目前比较成熟的 PKI / CA 体系的框架、技术和方法得到的。它与 PKI / CA 体系有很多相似之处：都要有管理对象，都要有公正的、权威的、可信赖的第三方作为管理者，并由此产生申请、审批、颁发、检测、撤销等诸多问题。但是两者又存在根本性的区别：PKI / CA 体系主要进行身份鉴别，即回答“你是谁”；数字水印基础设施主要进行版权确认，即回答“你是否拥有该电子作品的版权”。两者之间的对比可以从表 9.1 清楚地看出。

表 9.1 数字水印版权系统与 CA 体系的对比

概念	PKI 实体	数字水印实体
管理对象	证书	数字水印
管理者	认证权威	数字水印管理机构
用户	各种需要确认身份的主体	版权所有人
捆绑关系	证书将用户身份和版权要捆绑在一起	数字水印将用户、数字作品和用户对数字作品的版权捆绑在一起
信任源	根 CA	根管理机构
次级权威	子 CA	次级管理机构

9.1.3 系统实现方案

数字水印版权保护系统的实现是一个复杂的系统工程，决非单个人员在短时间内能够完成，它需要许多专家和技术人员的细致研究和共同努力。下面介绍具体的实现思路。

1. 角色定义

数字水印版权保护系统主要包括三个主体角色：

(1) 内容提供商（商家）

内容提供商在多媒体信息 X 中嵌入水印 W 。

(2) 认证第三方（服务器）

由三个模块组成，即内容提供商输入处理、用户输入处理以及电子证书分发。发生版权纠纷时，内容提供商的版权标识 W 可以从多媒体信息中提取。服务器发放证书，并产生电子证书解密密钥，发放给用户。

(3) 用户（客户）

用户购买产品后，注册个人信息；之后他将获得电子证书及其解密密钥。

2. 水印生成

在数字水印生成过程中，由客户提供标识、标记，水印中心生成水印信息，最后得到客户水印信息。为了证明自己身份，在以后纠纷出现时进行第三方公证，客户向数字水印认证机构请求一个合法的客户数字水印 W_B 。在确认客户身份后，认证中心随机为其生成数字水印 W_B ，用客户的公钥 K_B 对其加密后得 $E_B(W_B)$ ，连同认证机构的签名信息 $Sign_C(E_B(W_B))$ 发送给客户。

3. 水印嵌入

数字水印嵌入过程可描述如下。

(1) 客户把水印信息 $E_B(W_B)$ 和认证机构签名 $\text{Sign}_C(E_B(W_B))$ 发送给商家, 商家验证签名水印信息的合法性。

(2) 设 X 代表客户希望从商家那里购买的数字产品。由商家产生一个唯一标识商家的水印信息 W_A , 将其嵌入数字作品 X 中, 得到含水印作品 $X' = X + W_A$ 。

(3) 产生一个随机变换 Q , 用此随机变换作用于从客户处得到的水印信息, 得到 $Q(E_B(W_B)) = E_B(Q(W_B))$ 。

(4) 商家再把经过变换所得的水印信息 $Q(E_B(W_B))$ 作为第二个水印信息嵌入到已经嵌有水印的 X' 中, 得到 X'' 。然后, 商家用客户的公钥 K_B 对嵌有商家的水印信息的 X'' 进行加密, 得到 X''' 。

(5) 商家记录下客户的标识 ID_B 、唯一的水印信息 W_A 、从客户处获得的用客户公钥加密的水印信息 $E_B(W_B)$ 、经认证机构验证签名信息 $\text{Sign}_C(E_B(W_B))$ 和随机变换 Q , 把它们保存在一张表 Table 中, 以便将来纠纷调解使用。

(6) 客户收到来自商家的嵌有水印的作品 X''' 后, 用自己的私钥进行解密得到 X'' 。

现在客户就拥有了图像 X 的一个复制 X'' , 其中嵌有商家的水印信息, 即商家的版权信息, 这对其他任何人来说是不可知, 也不可能将它除掉, 这样商家标示了自己的版权所有。当中还嵌有客户的水印信息, 而且只有客户能够用其私钥对 X''' 解密得到带有水印的明文数据 X' 从而使得商家无法复制该复制; 另一方面, 出于随机变换对客户不可知, 客户也不能从中删除掉 W_B 。

4. 侵权判定

侵权判定协议用于防止盗版侵权, 追踪盗版起源。由于商家在自己的数字作品中加入了版权信息 W_A , 因此商家可以声明数字作品 X 的版权属于自己。一旦发现了 X 的非法复制 Y , 通过检测嵌入其中的水印信息 $E_B(W_B)$, 便可以确定复制源自哪位客户。

以 X 、 Y 作为输入参数, 水印提取函数可提取出复制中的客户水印信息 W , 在表 Table 中确定具有唯一水印信息的客户标识。在表中确定 W 的提取机制依靠所使用的数字水印技术。对于鲁棒水印, 通常把水印 W 与表 Table 中的每个水印信息进行相关性分析, 选择其中超过门限 T 且相关性最高者。如果 W 在表中找到, 商家可由客户 ID 获得客户的信息, 从而可以推断复制起源于该客户。如果 W 不存在于表 Table 中, 那么判定失败。

5. 侵权调解

一旦出现非法复制 Y , 如果客户否认自己是非法复制 Y 的起源, 商家可以提交 $E_B(W_B)$ 、 $\text{Sign}_C(E_B(W_B))$ 和 Q 给第三方仲裁者。仲裁者要求客户提供其私钥, 在 Y 中检测是否存在 $Q(W_B)$ 。如果存在则可决定客户是非法复制提供者, 否则客户是无辜的。这是一个三方协议, 当然如果客户将其私钥交给可信任的第三方, 客户也可以不参加在此协议之中。此时协议成为两方协议, 仍然可用于纠纷调解。

对上述这套协议, 需要说明的是, 由于商家变换了客户的水印信息, 尽管客户是唯一知道 W_B 的一方 (除了随机生成 W_B 的认证机构), 他也不可能从其图像复制中删去自己的 W_B 信息。商家能卖给客户的是唯有客户能解开的加密复制, 他不可能有相应的解密复制, 完全能排除商家制造含有客户标识明文复制的可能。一旦出现盗版纠纷, 商家可以提供前面论述中提到的证据, 由仲裁者决定非法复制的出现是否归咎于客户。

9.2 军事保密通信

保密通信是通信对象之间为防止机密信息被截取,按约定的方法改变信息的传输方式或表现形式以隐蔽其真实内容的通信方式。特点是对传输的信息在发送端进行加密变换处理或隐藏处理,在接收端进行脱密变换或提取操作恢复成原信息,使窃密者发现不了秘密信息的传输或者即使截获了传输的信号,也不了解信号所代表的信息内容或其中隐藏的信息内容。由于数字化语音通信广泛应用于社会的各个领域,人们越来越重视其安全性,这里介绍一种将信息隐藏和加密技术相结合的数字化语音保密通信系统^[156]。该系统将秘密语音加密后隐藏于公开的语音中,确保语音通信保密性和安全性。

9.2.1 系统组成

顾名思义,伪装式语音保密电话系统就是利用公共电话网即 PSTN 将信息隐藏和密码等技术进行有机结合的保密通信系统。系统选择应用较为普遍的 Modem 传输手段,主要原因是设备简单、易于购买、使用普及。该系统的通信双方只要各有一台 56kbit/s 的调制解调器即可,通信两端可以均是模拟调制解调器,传输协议采用 V.34,数据传输速率最高可达 28kbit/s。Modem 与 PC 的连接采用串口连接,并利用 Windows XP 系统提供中断方式驱动的串行通信驱动程序 COMM.DRV 进行数据接收和发送。该系统所需设备除 Modem 外还包括计算机语音采集卡、耳机和麦克风,如图 9.2 所示。该系统可实现双工伪装式保密电话通信,可像普通电话一样使用。系统运行时,首先进行 Modem 初始化然后拨打对方的电话号码进行通信连接,接着系统启动声卡采集秘密语音(采样率 8kHz、精度 16bit),然后利用选定的低速率语音编码算法(2.4kbit/s)进行秘密语音编码和加密处理,之后利用隐藏算法把秘密语音信息隐藏在事先准备好的另一段可以公开的用嵌入算法进行编码产生的混合语音码流(编码速率 13kbit/s)中,存放在发送缓存区内。发送线程将此部分数据读入 RCC(Reliable Comm Communication)内部发送缓冲区内,按 RCC 协议进行分帧、切割进行发送,通过 Modem 将此段带有秘密语音信息的语音编码数据传输给对方的 Modem。在接收端,由其 Modem 进行数据的解调接收,将帧数据中的帧头去掉,混合语音码流数据经检查无误后存放在接收缓冲区内。由提取算法提取在混合语音码流数据中隐藏的加密的秘密语音信息,解密算法对此数据进行解密,就得到秘密语音的编码数据,然后利用相应的低速率解码算法(2.4kbit/s)合成秘密语音并存放在声卡播放缓存区内,再由计算机声卡进行播放。

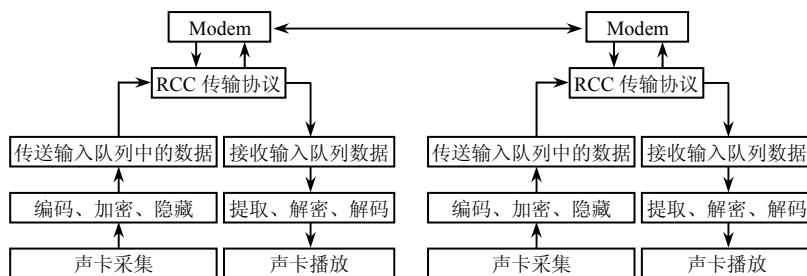


图 9.2 系统总构架

隐藏算法将秘密语音编码隐藏在公开的语音编码(即载体语音)中,携带秘密语音信息的混合码流按照正常 GSM 解码算法进行解码后的重构语音的音质良好,与原始载

体语音相比较, 虽然语音质量略有下降, 但在无对照的情况下, 完全可以认为是正常的通话, 从而使得“被动攻击者”难以觉察到所听到的语音中携带了秘密语音数据, 达到了隐藏秘密语音通信的目的。

9.2.2 系统采用的秘密语音隐藏方案

1. 总体方案

1982 年开始, 欧洲邮电管理会 (CEPT) 设立移动通信特别小组 (GSM) 开始研究制定泛欧数字蜂窝系统的标准。1985 年 10 月, 由英国、芬兰、德国、法国、意大利、荷兰、挪威及瑞典等国参加的语音编码专家小组从最初提交的 20 多种语音编码方案中优选 6 种进行了测试, 最后以 MPE-LTP 和 RPE-LTP 两种为蓝本, 制定了码速率为 13kbit/s 的, 带有长时预测环节的规则脉冲激励线性预测编码器 (RPE-LTP)。

系统选择 GSM 编码作为载体编码方案, 原因如下: 首先, GSM 编码方案具有比较低的码速率 (13kbit/s), 并且重构的合成语音的质量很好; 其次, 该语音编码中应用的部分语音参数具有比较强的鲁棒性, 这些参数的少量变动对于重构语音质量的影响较小, 这是选择该编码为载体语音编码方案的最关键原因。结合语音信息隐藏方法和 GSM 编码算法自身的特点, 可以构造以 GSM 语音编码为载体的大隐藏容量、运算简便的信息嵌入/提取算法。经过实际测试, 发现在 GSM 编码的数字码流中可以隐藏 3kbit/s 左右的语音信息, 并且嵌入后的混合编码重构语音质量也比较好, 这样就解决了实时嵌入秘密语音所需隐藏容量大的难点, 并且具有比较好的隐藏效果。因此, 利用该隐藏算法, 以 2.4kbit/s 的低速率语音编码 (MELP) 作为秘密语音编码方案, GSM 编码作为载体语音的编码方案, 将秘密语音码流按照嵌入算法实时嵌入载体语音码流之中, 构成利用 Modem 和计算机的基于信息隐藏的伪装式语音保密通信系统。

2. 嵌入过程

下面叙述秘密语音嵌入的具体过程。为叙述方便, 规定符号 T_1 为 GSM 编码一帧的时间长度, T_2 为选用的低速率编码一帧的时间长度, T 为嵌入算法的帧长 (缓冲区长度)。 $T=[T_1, T_2]$, 即 T 为 T_1 和 T_2 的最小公倍数, 为确保载体语音和秘密语音满足实时同步发送, 可确定嵌入算法延时 T 为 180ms。嵌入算法流程如图 9.3 所示, 具体过程描述如下。

(1) 预处理。发送方将载体语音先进行预处理, 去除直流分量并进行高频分量预加重。预加重采用一阶 FIR 滤波器, 目的是为了更好地进行 LPC 分析, 然后存储待用;

(2) 分帧处理。以每 20ms (160 个样点) 为一帧, 进行 LPC 分析;

(3) 短时分析滤波。这部分对信号 S 做 LPC 短时预测分析, 产生短时余量信号;

(4) 长时预测语音信号 S 经短时预测分析之后, 其余量信号 d 进入长时预测, 进一步去除信号的多余度;

(5) X 为原始秘密语音经过 A/D 采样后的 16bit 线性 PCM 码, 又经选用的低速率语音编码后产生的码流序列, X_i 为第 i 帧 48bit 秘密语音比特流;

(6) 将秘密语音 X_i 按照预定嵌入方法和规则激励码编码同步进行嵌入编码;

(7) 输出混合语音码流序列, 并将混合后的码流按照传输协议进行传输。

如此循环, 直至秘密语音全部嵌入后, 该编码过程恢复为正常 GSM 编码流程。

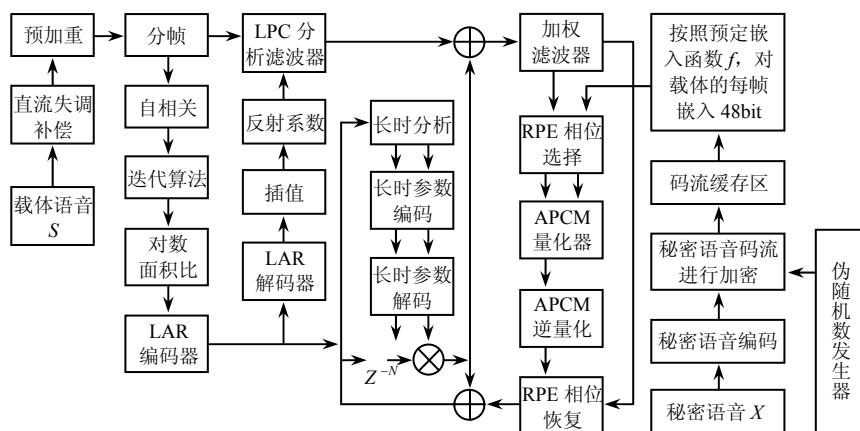


图 9.3 嵌入算法框图

3. 提取过程

提取算法流程如图 9.4 所示。接收方收到混合码流后，按照约定的分段长度对混合码流分段，对每一段的混合码流按照嵌入信息提取方法 f^{-1} 将秘密信息提取，并输入到选用的低速率解码器中，合成秘密语音输出到耳机。

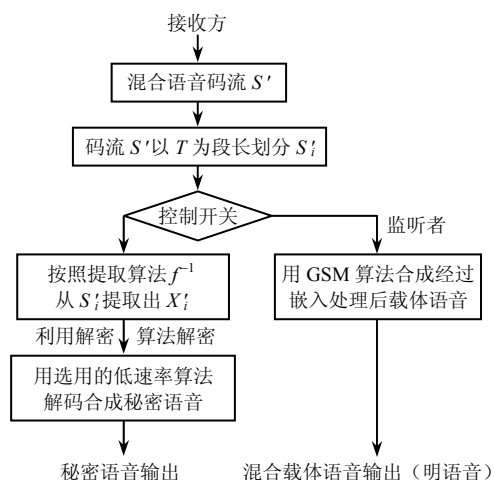


图 9.4 提取算法框图

9.2.3 隐藏效果

为了验证隐藏效果，下面给出一段载体语音的时间波形图和频谱图（图 9.5）以及嵌入秘密语音后重构合成语音的时间波形图和频谱图（图 9.6）的比较。从图形的对比情况来看，时域波形有一定差异，而两者频谱基本一致。经过实际试听，证实混合载体的合成语音的音质虽略有下降，但清晰度和可懂度基本不变，仅仅是噪声比原始语音大一点，较好地达到了隐藏秘密语音的目的。

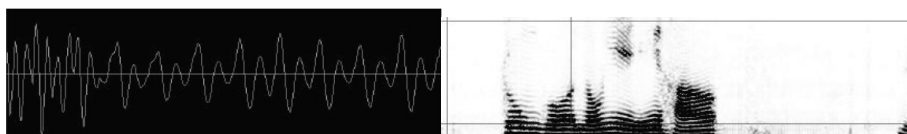


图 9.5 载体原始语音时间波形图和频谱图

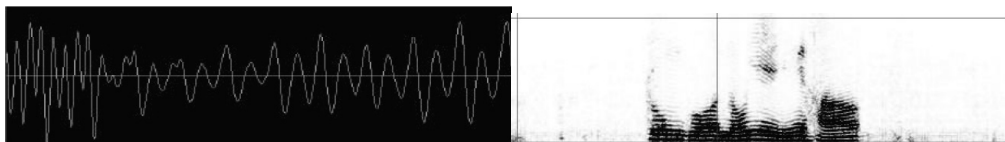


图 9.6 嵌入后载体重构合成语音时间波形图和频谱图

9.3 交易跟踪

数字指纹是解决交易跟踪的一种信息隐藏手段。数字指纹实际上是一种鲁棒的数字水印，它与用于版权保护的数字水印区别在于它在数字多媒体中隐藏的是购买者的信息，其目的是为某个购买者的一次购买过程提供信息证明，这与日常生活中所说的指纹有相似之处，都用来唯一性标识身份。当发现盗版作品时，可以提取其中的指纹信息，从而确定非法泄露者予以制裁。数字指纹解决了数字水印只能判定版权，不能进行盗版源头追踪的问题。数字指纹体制包含两个部分：一是用于向作品中嵌入指纹，并对带指纹作品进行分发的复制分发体制；二是实现对非法泄露者进行追踪并审判的追踪体制。

用户在购买作品的同时，会为用户分配一个唯一的指纹并嵌入到用户购买的作品中。用户获得嵌入指纹后的作品后，自己可以自由播放观看。某些非法用户可能会直接分发其购买的作品，或与其他用户联合生成新的复制后分发（即共谋）。无论是哪种情况，非法分发的复制中都会留下参与非法活动用户的指纹信息。一旦发行商发现了非法复制，就可运用指纹提取及指纹解码技术，追踪到非法分发者。下面以视频作品为例，介绍一种典型的交易跟踪应用方案^[157]。

9.3.1 系统整体架构

视频作品版权追踪系统总体目标是以数字指纹、复制检测技术为核心，紧密结合现有的数字版权管理（DRM）系统，构建网络环境下数字视频非法复制的发现、追踪、认定和监视框架。该系统以数字指纹、复制检测、密码学、DRM 等为支撑，实现视频版权主动追踪，克服了传统指纹系统的被动缺陷，具有很大的创新性。追踪系统主要提供版权注册与认定、指纹生成及颁发、指纹嵌入与检测、非法复制主动搜索与检测、泄露者的身份追踪与认定、版权动态监视等一体化服务，系统总体框架如图 9.7 所示，其基本原理如下。

（1）版权注册服务器和指纹生成与管理服务器部署在第三方权威机构（如版权局），主要为内容所有者提供版权注册登记服务，便于以后的版权认定；指纹生成与管理服务器为消费者提供指纹生成、颁发服务；

（2）内容所有者注册自己的版权后，可通过 DRM 系统对其作品进行加密打包，然后通过内容分发/销售网络分发到消费者客户端，消费者可通过 DRM 系统购买许可证；

（3）指纹生成与管理服务器提供用户指纹证书的颁发和验证服务，消费者首先通过申请，获得指纹证书，在通过 DRM 系统购买许可证时，需要提交指纹证书到 DRM 版权中心。DRM 版权中心对该指纹证书进行验证，只有验证通过后才能发送内容许可证；

（4）合法消费者得到内容许可证后，客户端可以对加密内容进行解密，然后利用指纹嵌入器嵌入用户指纹，将内容输出到播放器/解码器供用户播放/浏览；

（5）合法消费者在内容播放/解码过程中，可能利用录制软件、刻录光驱、摄像机等将内容复制下来，然后散布出去，可能在用户群内传播，也可能在用户群外的网络传播；

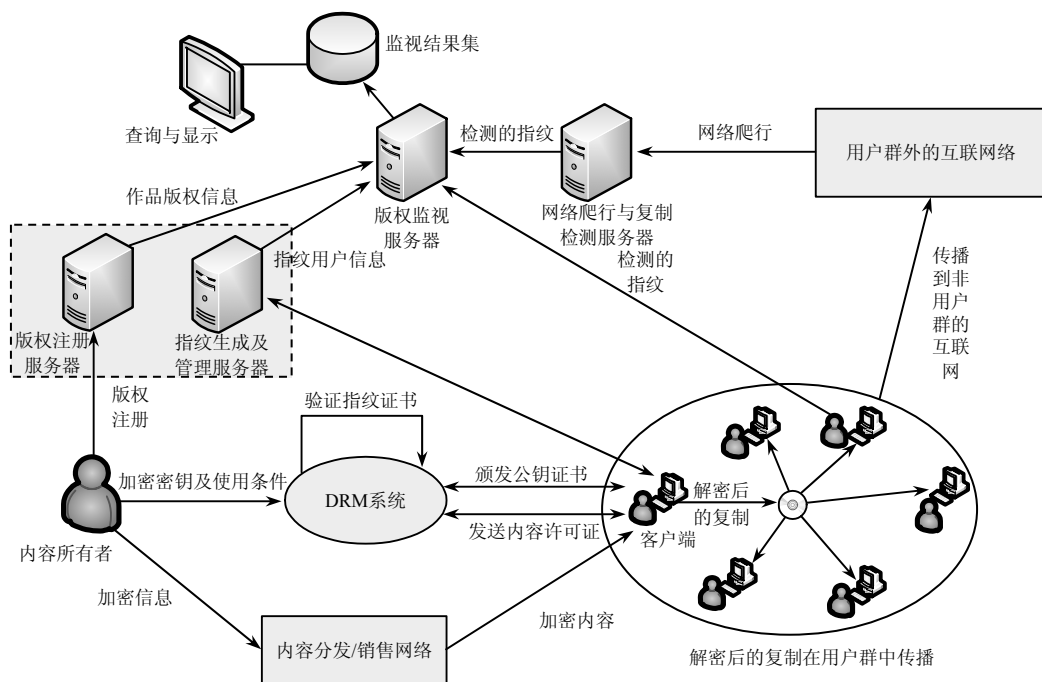


图 9.7 版权主动追踪与监视系统总体框架

(6) 如果在用户群内传播，客户端通过动态扫描和指纹检测，将监视的结果（指纹、时间、IP 等）发送到监视服务器；

(7) 对于用户群外的网络，采用网络爬行和复制检测服务器得到非法复制作品，然后利用指纹检测器，将检测的结果发送到监视服务器；

(8) 监视服务器通过第三方权威机构的指纹信息数据库查询，得到非法复制者的信息，而版权的认定可以通过第三方权威机构的版权注册信息加以判断；

(9) 内容所有者可以通过终端与监视服务器相连，随时关注自己的作品是否被非法复制，以及被哪个消费者复制的。

据此，将该系统划分为原始作品数据库、特征数据库、监视结果数据库、网络媒体数据库、数字指纹数据库；版权注册管理模块、数字版权管理模块、数字指纹检测模块、客户端模块、指纹证书管理模块、版权监视模块、复制检测模块、网络爬行模块、用户查询与显示接口等几个核心部分。

原始作品数据库用于存储版权所有者的数字视频信息，以作为版权认定的凭证；特征数据库用于存储数字作品 ID 以及经过版权注册管理模块处理以后得到的数字作品的特征码；监视结果数据库用于存储版权监视结果，包括作品 ID、非法使用该作品的用户 ID 以及监视时间等；网络媒体数据库用于存储从网络上爬行下来的数字作品信息；数字指纹数据库用于存储用户 ID 以及对应的数字指纹。

版权注册管理模块用于向版权所有者的提供版权注册服务，以服务器的形式部署在第三方权威机构，其主要接收来自版权所有者的数字作品文件，然后从数字作品中提取特征码，并将提取到的特征码存储到特征数据库，完成数字作品的注册；数字版权管理模块接收从客户端模块发来的数字指纹证书，并对数字指纹证书进行验证，验证通过以后向该客户端模块发送内容许可证，否则不发送；数字指纹检测模块检测网络爬行模块或客户端模块提供的数字作品中的数字指纹，并与数字指纹数据库中的用户指纹进行相

关性检测, 追踪到传播该数字作品的非法用户将检测结果发送到版权监视模块; 各客户端均设置有客户端模块, 客户端模块向指纹证书管理模块申请数字指纹证书, 发送数字指纹证书到数字版权管理模块, 并获得内容许可证, 解析内容许可证中的解密密钥, 实现对数字内容的解密, 同时向数字作品中嵌入数字指纹; 指纹证书管理模块以服务器的形式部署在第三方权威机构, 用于响应客户端模块的数字指纹证书申请请求, 然后生成数字指纹证书并颁发给客户端模块; 版权监视模块用于接收数字指纹检测模块发送过来的版权监视结果, 并将监视结果保存到监视结果数据库; 复制检测模块用于提取网络爬行模块从互联网上爬行到数字作品的特征码, 并利用特征码匹配算法找出该数字作品的复制; 网络爬行模块用于爬行互联网上的可疑数字作品, 并得到数字作品的相关信息, 然后调用复制检测模块, 判断该作品是否是已注册版权的数字作品的复制版本, 如果是复制, 则调用数字指纹检测模块对其进行指纹检测, 如果不是复制, 则不做处理; 用户查询与显示接口用于接收版权所有者对其已注册版权作品的查询请求, 然后查询监视结果数据库, 并将查询结果显示给版权所有者。

9.3.2 版权注册与认定

版权注册与认定非常重要, 可以有效解决版权纠纷问题。本系统采用多媒体自身特征作为零水印, 同时借鉴公钥认证体系相关技术, 制定完备的版权认定协议和流程实现数字视频作品的版权注册与认定。

1. 版权注册

版权注册流程如下。

(1) 利用复制检测技术检查作品或其复制是否已注册这里采用复制检测, 主要是为了防止重复作品注册

如果用户拿一个已注册视频作品的复制来进行注册, 通过复制检测技术, 就很容易判别出它是否是原始作品的复制, 防止重复注册。如果该作品已注册, 则拒绝版权注册, 否则上传数字作品。这样可以保证版权的唯一性。

(2) 从申请注册作品中提取全局特征

对作品进行认定, 要求准确度高、速度快, 不需要抵抗一些复杂的处理和攻击。只要原始作品经过修改以后就可认定为不是同一个作品, 从而判定当前认定者不是作品所有者。研究发现, 由分块 DCT 系数构成的顺序测度特征具有较高的辨识能力, 且计算速度快, 因此系统采用分块 DCT 特征。

(3) 使用哈希函数进行哈希运算, 将特征信息映射成固定长度的哈希特征码

数字视频作品的特征信息量较大, 并且不同作品的特征信息长短不一。为了便于检索和存储, 对其进行哈希运算, 将特征信息映射成固定长度的哈希特征码。

(4) 版权注册处理

利用第三方权威机构的私钥对哈希特征码进行签名, 并与注册者信息关联, 保存在原始作品数据库中。由于签名具有不可伪造性, 因此可以保证注册信息的权威性。

(5) 版权公示及转正

在版权公示期间, 如果无版权争议, 则进入版权转正。如果存在争议, 则注销作品信息。版权转正后, 会更新特征索引数据库, 将作品特征加入索引。由此, 内容所有者的作品正式注册成功, 版权由此刻起开始生效。

2. 版权认定

视频版权认定过程支持在线和离线两种模式。对于在线认定过程,因为服务器查找版权所有人只需要作品的全局特征,因此不必将作品上传,而是将客户端得到的哈希特征码上传到版权注册服务器。这样就节省了网络带宽,也减轻了服务器负担。在线认定时,申请者先提交作品,客户端自动提取作品的全局特征并进行哈希运算得到哈希特征码。然后将该特征码上传到版权注册服务器,并采用与版权注册相同的方法对特征码进行私钥签名,最后对数据库中已注册作品的哈希特征码进行检索,借助于关联的注册者信息,可输出谁是版权所有人。对于离线认定过程,版权注册成功后会为注册用户提供作品证书的下载服务,证书内容包括:签名的哈希特征码、所有者信息和有效期等。离线认定时,用户提交作品证书和数字作品,客户端首先验证证书的有效性,如果证书有效就将特征码解密;然后提取作品的全局特征、计算哈希特征码,并与证书解密的特征码相比较,判断是否一致;最后可得出该证书的持有者是否是版权拥有者。

9.3.3 指纹证书管理

指纹生成与管理服务器负责指纹的生成、指纹证书的颁发和验证,其部署在第三方权威机构,一方面保证交易协议的非对称和匿名性,另一方面可防范内容提供商与版权中心共谋,实施陷害消费者而获取额外利益的行为。以下将从指纹证书生成、指纹证书颁发协议和指纹证书的验证协议三个方面来详细介绍。

1. 指纹证书生成

数字证书的格式一般采用 X.509 国际标准,主要包含如下内容:版本号,指出该证书使用了哪种版本的 X.509 标准;序列号,证书的唯一数字型编号;签名算法,用来指定 CA 签发证书时所使用的签名算法;颁发者,颁发该证书的机构 CA 的名字;有效期起始日期,该证书可用的开始日期;有效期终止日期,该证书可以使用的截止日期;主题信息,证书持有人唯一的标识符(Distinguished Name, DN);公钥,和证书持有人私钥相对应的公钥及其他相关参数;扩展信息,在不改变证书格式的前提下附加标识信息,包含证书的约束,密钥用途等内容。本系统将加密后的指纹就添加在扩展项中。

由于数字指纹都是由 $(-1,1)$ 范围内的浮点数构成,为方便加密处理,先将其转换成字符形式。具体操作是,首先将每个浮点数乘以 10000,丢掉小数部分,则可以用 16bit 表示;将这 16bit 的高 8 位和低 8 位分别用一个字符表示,则对长度为 L_w 的指纹,转化成长度为 $2*L_w$ 的字符数组。这样就完成了从浮点数到字符的转换。但这还有一个问题,即某些转化出来的字符超过 ASCII 码字符的 $[0, 127]$ 的范围,直接写入证书会出问题,于是采用 BASE64 算法对字符串再做一次编码,将所有字符映射到 64 个可见字符。

2. 指纹证书颁发

用户首先向第三方权威机构提出申请,审批合格后,由指纹生成与管理服务器分配指纹,然后生成指纹证书,颁发给用户。用户指纹证书与公钥证书结构类似,包含指纹密文与 Hash 签名、权威机构公钥和有效期等。指纹密文是以用户公钥加密指纹得到的,然后计算密文的 Hash 序列,再对 Hash 序列利用权威机构的私钥签名。指纹证书颁发的具体过程如下:首先,用户注册成功以后,输入其 ID 登陆到指纹证书颁发页面,系统根据用户 ID 从数字指纹数据库中找到分配给该用户相应的指纹 W ;其次,用该用户的公钥对指纹进行加密得到指纹密文;然后,利用权威机构私钥签名指纹密文得到签名指

纹序列；最后，将用户信息数据库中的用户信息、得到的签名序列以及加密后的指纹一并写入数字指纹证书中，得到最终颁发给用户的数字指纹证书。

3. 指纹证书验证

版权管理服务器在发放许可证之前需要对用户指纹证书进行验证。由于采用权威机构的私钥签名，在私钥未知的情况下指纹证书很难伪造。验证过程中，都是验证指纹的密文，版权中心无法知道用户的指纹，可有效地防止内容提供商与版权中心合谋。指纹证书验证的具体过程如下：首先，读取指纹证书，并验证证书有效期，超过有效期则证书无效；其次，如果证书在有效期内，则提取证书中的指纹密文和权威机构的私钥签名；然后，对指纹密文以及指纹签名利用相同的哈希函数进行哈希，得到指纹 Hash 值和签名的 Hash 值；最后，比较指纹 Hash 值和签名 Hash 值，如果两者相同则通过验证，否则证书无效。

9.3.4 指纹嵌入与作品分发

视频作品指纹嵌入在客户端进行，首先需要获得加密的视频的作品，然后再嵌入消费者指纹信息。

1. 加密作品分发

由于视频的容量很大，直接通过点对点的传输，一旦网络出现故障，控制起来会非常复杂。一种解决方法是将作品加密后挂到服务器上，供消费者下载。这种方式降低了客户端对服务端的依赖性，但加密算法的安全性至关重要。

本系统根据消费者在购买作品时提供的指纹证书，提取其中加密的指纹，进行一次哈希加密，以此作为加密密钥，对视频作品进行加密。由于数字指纹证书已由第三方权威机构的私钥签名，可保证不可伪造性，因此，通过此种方式加密的视频作品，只有该消费者才能解密，从而保证安全性。

另一方面，视频作品挂载到服务器上，对服务器的磁盘空间有一定要求，因此需要对此提出管理策略。本系统对每一次购买行为制定一个下载截止日期，过了这个日期后，会自动删除此次购买生成的加密视频作品。

2. 作品指纹嵌入

客户端在从指纹证书管理服务器获得有效指纹后，提交到版权管理模块进行验证。通过验证后，版权所有者会生成一个加密的视频作品供消费者下载，并颁发内容许可证。在消费者解密视频作品的同时，会从指纹证书中提取消费者指纹，嵌入到视频作品中，嵌入算法采用第 2 章描述的算法。

由于加密密钥是由消费者的指纹证书哈希运算再加密而来，则解密时，只需用同样的算法即可得到解密密钥。

9.3.5 作品版权主动追踪

消费者得到嵌入指纹的视频作品后，可能会将该作品上传到互联网。本系统利用网络爬行技术，对互联网上的视频作品进行主动搜索，并利用复制检测技术判断是否是版权作品的复制，如果是，则提取其中的非法用户的指纹，追踪到非法用户。指纹检测追踪具体步骤如下。

- (1) 利用复制检测技术从原始注册视频库中找到与可疑复制视频对应的原始视频;
- (2) 提取复制视频中的消费者指纹。
- (3) 对指纹进行 RS 译码追踪, 得到用户唯一标识串, 查找用户数据库信息, 将监视结果发送到监视服务器。

9.4 真伪鉴别

真伪鉴别也叫内容认证。对内容的认证包含两个层次, 第一层次是做出是否发生篡改的判断, 第二层次是能对篡改的部位或者可能遭受的篡改操作做出估计和判断。这类应用大多采用脆弱水印, 但也有利用鲁棒水印判断篡改的成功例子。这类水印对于新闻图片、案件取证图像、医学图像和军事图像等具有重要意义。这类水印也能在电子商务的交易中发挥作用, 例如检验电子票据、电子印章的真实性和完整性。下面以公章图像认证为例, 介绍一种典型真伪鉴别应用方案^[158]。

9.4.1 电子公文特点

纸质公文的安全性通过公章来保证, 一份盖有公章的纸质文件具有法律效力; 电子环境下公文的安全性通过含有脆弱水印的公章来保证。开发公章中的脆弱水印算法需要对电子公文及公章进行分析, 确定公章的使用限制, 从而设计一个具体的脆弱水印算法。

1. 电子公文的产生及特点

公文就是在公务活动中形成的文字材料, 一般是指党政机关、群众团体和企事业单位在处理公务时使用的有一定规范的各种正式文书。在信息化的今天, 各地政府正在大力推行无纸化、电子化、自动化办公模式, 相当数量的公文正在或将要以电子文本的形式传送, 出现了电子公文的概念, 电子公文是将纸质公文 (特别是其红头、公章和签名等) 运用计算机系统和现代信息管理技术制发的全数字化形式的公文, 如图 9.8 所示。

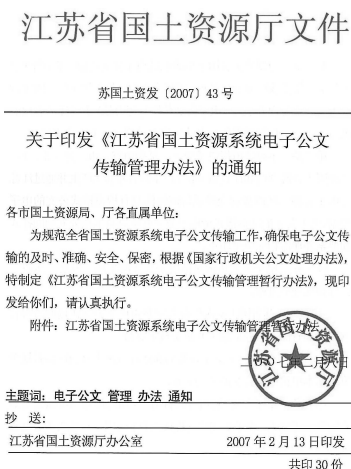


图 9.8 电子公文示例

纸质公文上的信息 (红头、正文内容、签名、公章等) 被固定在纸质载体本身上, 信息与载体构成不可分离的完整实体, 因而具有原始性特征。电子公文的信息与载体是分离的, 电子公文本身作为一个电子文件, 是以与存储实体分离的非实体的形态而存

在、被处理的,这种分离特征使电子公文具有可更改性、可复制性的特点。随着办公自动化的发展,电子公文交互系统的实施,改变了人们现有的办公方式,然而现有的公文交互系统还不能对收到的公文身份的真伪进行鉴别,公文就失去了法律效力。

2. 公章图像的使用限制

电子公章图像通常是经扫描仪扫描且经过处理后的 RGB 图像,具有如下使用限制。

- (1) 公章颜色比较简单,理论上仅有两种颜色,所以水印只能嵌入到红色图案中。
- (2) 为了达到纸质公章同样的效果,公章需经过镂空处理,因而可嵌入的信息量少。
- (3) 考虑到公文接收者对处理时间的忍耐程度,公文处理需要速度快,因而算法要处理简单、易于实现。
- (4) 攻击者的主要目的是攻击公文的内容,水印检测时可忽略局部检测性能的要求。

9.4.2 嵌有水印公章的特点和需求

嵌有水印的公章应当具有与实物公章相同的功能,即保证公文的真实性和有效性,体现公章的法律效力,下面从嵌有水印公章的特点和需求分别阐述。

1. 嵌有水印公章的特点

通过对电子公文中公章的功能和特点分析,含脆弱水印的电子公章应具有如下特点。

- (1) 公章可被确认,即当文件上有公章时,别人确信这个文件是经盖章者发出的。
- (2) 公章是无法伪造的,即公章是盖章者的凭证。
- (3) 公章是无法被重复利用,即任何人无法把盖章者在别处的印章挪到该文件。
- (4) 文件被盖章后是无法被篡改的。
- (5) 公章具有不可否认性,即盖章者无法否认自己在文件上的盖章行为。

2. 嵌有水印公章的需求

电子公章应具有保障电子公文的安全性,即如何保障公文的合法性、可追溯性、防非法篡改以及电子公章的权威性等问题成为亟待解决的问题。如果将传统的公章系统毫不改变地移植到电子环境中,我们不难发现,会出现两个问题:一是即使公章图像难以伪造(通常也很难办到),但是可以从一个文件到另一个文件通过复制和粘贴这样方式来得到公章图像;二是文件在这样的盖章后也很容易修改,并且不会留下痕迹。这两个问题导致了电子公章不满足所要求的基本特征,可以看到,在这样设定下的电子公文是有两个基本要素组成的:文件内容和公章图像。这比传统公章下的签名要少一个要素:纸张。在传统的公章中,纸张将文件内容与公章印文有机地联系在一起,实现了签名与文件内容的相关性,因此,在电子化条件下的公章,也要实现签名与文件内容的相关性,既是说公章图像中必须要包含文件内容的信息,并且这样的信息因文件内容的不同而不同,因此,采用带有公文内容相关信息的脆弱水印的电子公章,就能够既满足纸质公文的流程要求,又能够解决接收公文的身份鉴别问题。

数字水印可以克服当前数字签名技术只能应用到电子文档,对于普通文件则无能为力的缺点,另外,数字水印可以在原有图像上隐藏多种信息,而不是把签名作为一个尾巴附加在文档后面。数字水印技术把数字签名无缝地融合到数字印章中,从而使得数字签名以数字或者纸质形态存在。水印算法识别被嵌入到保护对象的所有者的有关信息(如注册的用户号码、产品标志或者有意义的文字等)并能在需要的时候将其提取出来,用来判别对象是否受到保护。

9.4.3 公章认证方案

为了防止伪认证攻击,以及公章本身的约束条件,方案结合了以下几种策略增强其安全性,避免了已有方案的安全缺陷。

1. 与数字签名相结合

现有的用于图像认证的方法有密码学中的数字签名技术和信息隐藏中的脆弱数字水印技术。传统的基于密码学的数字签名技术也可以保护数据的完整性和真实性,随着PKI(公钥基础设施)技术的发展与普及,将电子文档利用密钥加密成密文后发布,在网络传递过程中使得非法攻击者无法从密文获得机密信息,可以在一定程度上达到文档信息不被非法获取的目的,但这并不能完全解决问题。一方面,加密后的文件因其不可理解性妨碍电子文档的传播;另一方面,电子文档经过加密后容易引起攻击者的好奇和注意,并有被破解的可能性,而且一旦被破解其内容就完全透明了。

使用水印技术有两方面潜在优势:第一,水印不需存储相关的超数据(如签名);第二,水印和含水印作品一起经历相同变换,与附加签名不同,当作品被改写时水印自身也发生变化。通过水印和已知变换的比较,不仅可知道发生改变,并且可推断出何时何地发生怎样的改变。密码术是以不可读的乱码形式出现的,而脆弱数字水印能够保持数字图像的直观显现,因此两者相结合,提供了一种很好的电子公文安全性解决方案。

2. 与图像内容相结合

目前的水印认证算法中,水印信息的形成大多与图像内容无关,若采用基于图像内容的水印信息,则一方面可以增强系统抵御攻击的能力,另一方面又可避免在认证检测时额外提供原始水印信息;而且图像内容一般对各类常用的信息处理操作具有鲁棒性,对恶意篡改具有敏感性,所以,这类依赖于图像内容的水印算法更适用于图像认证。这里,水印采用电子文档和电子公章图像的数字签名,因此水印算法具有与载体图像的内容相关性。

3. 空域中基于数的随机排列

公章作为一种特殊载体,进行镂空处理,水印仅能嵌在公章的红色部位,有规律地嵌入信息就会使攻击者很容易获取水印的嵌入位置,从而提取出水印,公文安全性不能得到保证,所以需打乱水印的嵌入位置。这里用一个基于数的随机排列的水印算法,水印的嵌入位置是随机的,不知道密钥的攻击者无法获得水印信息。频域方案虽然可见性好,但处理速度慢,能量分布到整个图形空间,不适用于公章,因此采用基于空域的脆弱水印方案。

4. 密钥一次一密

本方案用随机数控制水印嵌入位置,将公文内容的哈希函数的二进制值作为随机数种子的一部分,这样不同内容的公文的哈希值是不同的,可做到一次一密,安全性得到提高。

9.4.4 电子印章系统

电子印章(公章、图章等)系统是计算机技术在传统印章电子化过程中的应用,电子印章系统的设计,不仅体现了现代计算机技术对传统公文和印章的影响;同时,中国传统文化对计算机技术的渗透也清晰可见。在设计电子印章系统的过程中,不仅要考虑电子公文本身的安全性,还要照顾用户使用印章的传统习惯,如印章图片等。已有的采用数字签名技术来实现电子公文的身份认证,可是没有考虑到其安全性缺陷。电子印章

系统是基于当今流行的 Word 文档，实现在 Word 文档中的电子盖章，该电子印章系统是基于脆弱水印、数字签名及 Word 二次开发等技术，解决电子公文的身份认证和电子公章的权威性等问题。

1. 系统模型

为解决电子公文交互系统的安全性问题，所采用的电子印章系统模型如图 9.9 所示。

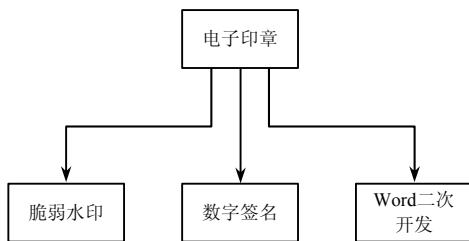


图 9.9 电子印章系统模型

电子印章系统具体内容如下。

(1) 采用用于公章图像的安全脆弱水印技术，保障电子公文的真实性、公章的权威性，并且符合中国传统习惯的公章行文要求。

(2) 数字签名采用 RSA 数字签名和 MD5 哈希函数，保障盖章公文的信息完整性。

(3) 在 Word 文档中实现电子盖章，即盖章过程与 Word 文档的整合，也就是 Word 二次开发技术。

2. 系统的设计思想

(1) COM 组件化设计思想

COM (Component Object Model) 是一种基于二进制通用接口的一组规范，并提供一系列的接口给外界调用。COM 组件是以 Win32 动态链接库 (DLL) 或可执行文件 (EXE) 的形式发布的，它具有语言无关性、进程透明性、可重用性、保密性等诸多优点。使用 COM 的发布方案，客户可以动态地找到他所需要的组件。OLE (Object Linking and Embedding) 技术以 COM 规范为基础，OLE 技术允许将其他应用程序 (例如电子表格、图形工具以及其他数据库) 的数据集成到自己的应用程序中。Active X 控件的主要技术基础为 OLE 技术，它几乎涉及了容器型应用程序与服务型应用程序之间交互的所有技术，并且 Active X 控件还引入了一些新的技术规范。除了属性和方法外，Active X 控件还包含事件，当某些动作发生时通过事件通知容器，并且易于嵌入到 Web 应用中。

(2) Word 的二次开发

由于 Lotus 系统没有提供强大的编辑功能，这就决定了其文字编辑、处理能力没有用户日常使用的 Word 的功能强大，所以系统调用 Word 进行公文处理。Word 本身是一个实现了自动化的 COM 组件，他提供了一个叫做对象库的文件，或叫类型库，其中包含了对象、属性、方法、事件等，可以用其提供的接口来访问和操纵 Word 文档，使其为自己编写的客户程序服务。

3. 系统的工作流程

采用电子印章系统作为一个控件设计的电子公文系统必须要灵活、高效、具有很强的可扩展性。一种可行的工作流程如图 9.10 所示。在发文端，用户通过身份认证登录到

电子公文交互系统中,新建公文,启用 Word 编辑软件,进行公文编辑,编辑结束后进行盖章申请,公章管理者通过身份验证从公章库中选取公章,分别计算公文和公章的哈希函数,公章管理者以其私钥对公文进行签名,根据水印的嵌入算法嵌入到公章中,将公文加密传输。在收文端,用户以合法身份登陆系统,进入收文管理模块,当接收公文时,将公文解密,打开公文点击验证按钮,从接收到的公文的公章中提取出水印信息,使用发方用户的公钥对接收到的公文进行解密,并分别计算公文文档和公章的哈希函数,将其值同解密信息进行比较,根据比较结果判断公文是否改动或印章是否为复制的,并进行相应的处理,显示提示信息。图 9.11 给出了某公司开发的电子印章系统界面示意图。

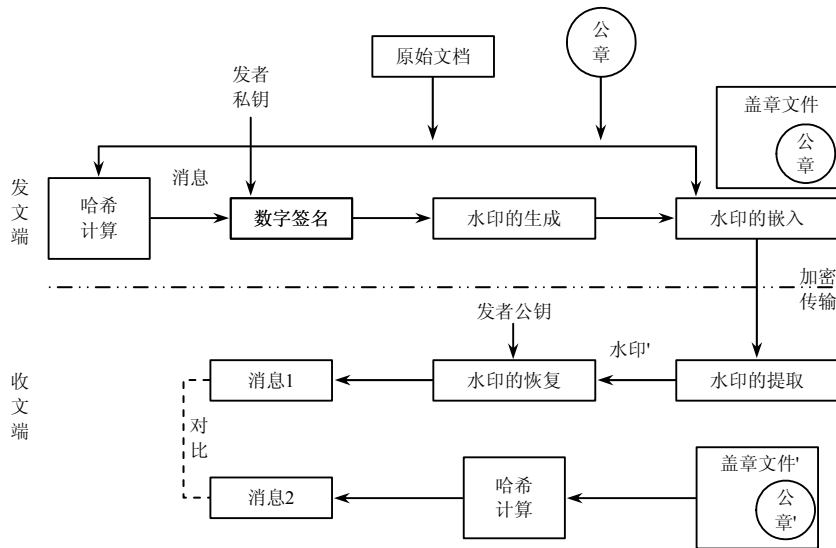


图 9.10 电子印章工作流程

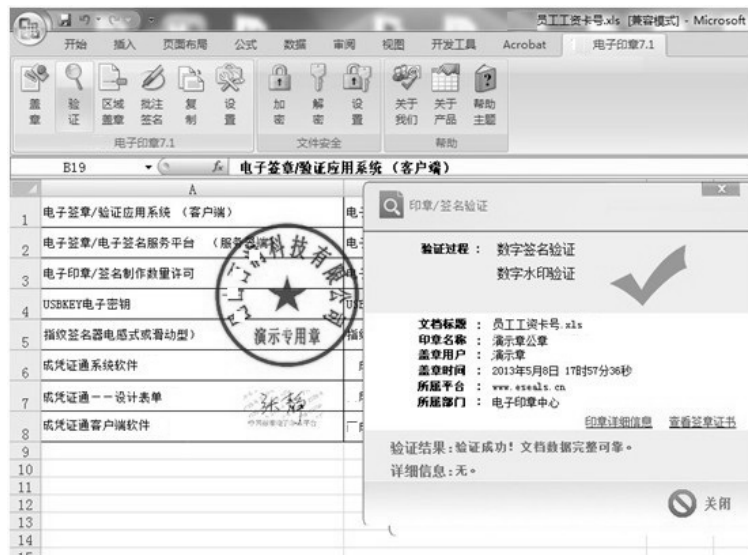


图 9.11 某公司开发的电子印章系统界面

4. 密钥的管理

在企业内部网中,公钥可由企业的主服务器和系统管理员来分配、管理和储存。在

广域网中，可由专门的发证机构来进行分配和管理。这里，每位员工可根据自己的 ID 号、登录密码和一些私人信息来向本系统申请一密钥对，并输入一使用密码。用户输入的签名认证密码即起到确认用户身份，据此来调用系统中保存的其私钥。这样，用户可以通过经常更新自己的密钥来提高其签名安全性而不增加额外的开销。

9.5 复制控制

复制控制技术指的是通过加密、水印等技术手段，限制将物理介质上的媒体内容复制到别的存储设备，限制非法复制内容，保护内容提供商的利益。本节以 DVD 复制控制为例，介绍水印的应用。1996 年，美国电影协会、消费电子产品制造商协会和部分计算机厂商联合成立了国际版权保护技术工作组（CPTWG）来研究防止数字视频，特别是 DVD 产品被私自复制的技术问题。CPTWG 在过去的几年中已成功地研制出 DVD 防复制系统的主要部分，而且这种 DVD 防复制系统将注定成为 DVD 事实上的标准之一。虽然 CPTWG 无权要求用户端设备必须采取这种版权保护措施，但由于被保护的 DVD 视盘必须进行加扰，所以不符合“标准”的设备是无法正常播放其内容的。下面首先介绍现有的复制控制技术，然后介绍视频水印技术用于复制控制需解决的问题，最后介绍一种典型的 DVD 复制控制系统方案。

9.5.1 现有的复制控制技术

CPTWG 系统不断发展，目前已有 3 种技术应用于其中，即：内容加扰系统（CSS），模拟信号防护系统（APS）和复制管理系统（CGMS）。同时，CPTWG 正考虑其他两种措施：在 PC 总线上进行秘密通信（由 5 家厂商联合研制，称为 5C）和水印技术。

（1）CSS 是三菱公司研制的对 MPEG2 视频流进行加扰的措施。它是一种较复杂的密钥技术。密钥分别存储在 DVD 视盘的导入区和节目扇段的头部，解码时通过专用芯片对密钥进行解码之后再对视频流内容进行解扰。CSS 的优点在于：密钥信息只能被专用芯片所读取，对于那些不具备专用芯片的播放设备将无法读取密钥信息，也就无法对视频流进行解扰；同时密钥信息也不可能被复制，确保盗版光盘不能被任何设备播放。

（2）APS 是美国 Macrovision 公司开发的一种对模拟电视信号进行改造，从而防止光盘上的数字化影音信号被转录为 VHS 模拟信号的新型技术。它在电视信号中加入伪水平同步脉冲，从而影响 VCR 的自动增益控制系统（AGC），使录像机录下的图像呈忽亮忽暗的变化，无法观看。但由于 TV 对 AGC 反应较慢，故仍能正常播放光盘机输出的电视图像。

（3）CGMS 技术仅在 MPEG 视频流头部加入几个 bit，用来标识三种状态：“自由复制”，“禁止复制”和“一次复制”。

（4）5C 是提供安全保障的通信系统。它可使符合 CPTWG 建议的设备通过计算机总线交换密钥以互相传输加密数据，而防止其他设备对秘密信息的非法解密。5C 系统是随着计算机高速总线（如 IEEE1394）的出现而发展起来的，它的潜在使用场合是将未经压缩的视频信息由视盘放像机或机顶盒向监视器传送。

假定市场上有三种类型的 DVD 盘片，如图 9.12 所示。经 CSS 加扰的合法 DVD 产品只能在符合“标准”的设备上播放。由于不含解扰密钥，经 CSS 加扰的盗版 DVD 盘片不能在任何光盘播放机上播出。但那些经过解扰的非法复制的盘片却可以在任何设备上播放，如图中的 1 所示，目前市场上这种盘片比例较大。

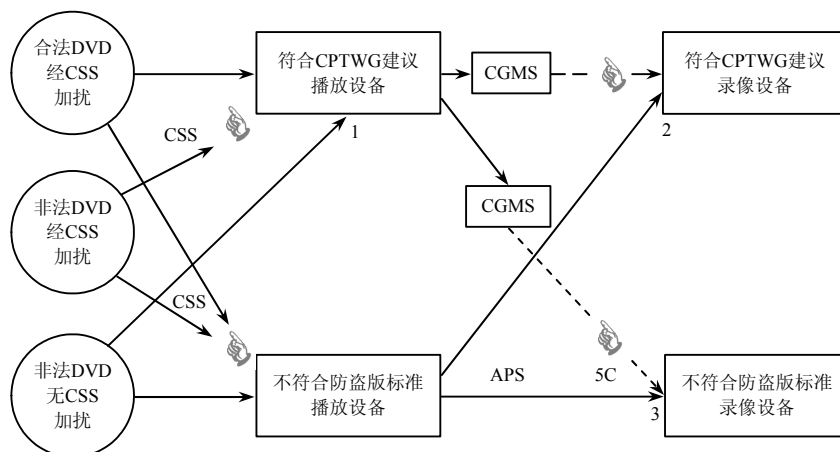


图 9.12 没有应用水印技术的 DVD 防复制系统

CGMS 系统用来控制盘片的复制次数。但非标准的 DVD 播放设备可轻易去掉 MPEG 视频流头部中的这些比特位，使复制盘片不再受到任何限制。这时完全可以利用各种翻录设备生产出没有 CSS 或 CGMS 保护的 DVD-RAM，如图中的 2 所示。APS 可防止将视盘节目翻录到 VHS 录像带上，而 5C 系统确保视频流只能在标准数字显示器上显示。它们目前的应用领域还比较窄，若播放设备输出的为模拟 RGB 信号，那么盗版者完全可以利用合适的翻录设备制造出不含任何保护措施的 DVD 盘片，如图中的 3 所示。综上所述，目前的 DVD 防复制系统还存在一些缺陷。该缺陷会导致大量盗版产品的出现，因为非法复制的盘片可被任何用户设备播放、翻录。而应用水印技术可弥补上述缺陷，如图 9.13 所示。

两类应用水印技术的模块被加入到 DVD 防复制系统中，分别是记录控制与回放控制。记录控制取代了 CGMS 的功能。它利用水印的鲁棒性将 CGMS 数据保护起来，保证复制控制比特不会被轻易除去，从而有效防止因消除有关数据而引起的非法复制。引入回放控制的优点在于：如盗版者成功地生成了不含 CSS 密钥信息的非法 DVD-RAM 复制，由于水印仍存于这一复制中，符合标准的光盘播放机将会读出受水印保护的复制控制信息并根据 RAM 盘片本身的特点做出拒绝回放的判断。这就将这种非法盘片的市场限制在拥有非标准播放设备的用户中，而另一方面，这种设备却不能播放合法的正版 DVD 光盘。

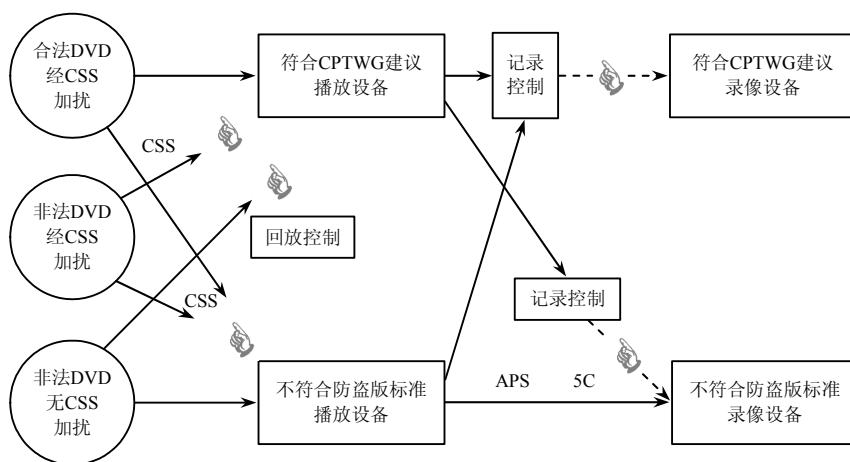


图 9.13 应用水印技术的 DVD 防复制系统

9.5.2 视频水印技术用于复制控制需解决的问题

1997 年夏天, CPTWG 专门成立了数据隐藏子工作组 (DHSWG) 来评价当前的水印技术应用于防复制系统的先进性和可靠性。有十几家公司向 DHSWG 提交了它们的方案。目前水印技术发展中还存在着下面的一些主要问题。

(1) 提高数字水印的鲁棒性, 即保障水印在经历各种非法攻击后仍能有效存在而不会被轻易消除是这项技术发展中的最关键问题。非法盗版者会想方设法除去正版光盘中的水印以达到复制盗版的目的。对于那种独立于图像内容的水印, 盗版者可通过帧平均的方法重构水印, 或通过所谓灵敏度分析、用图像质量退化的代价来重构一帧中的水印, 进而达到从视频序列中除去水印的目的。面对以上问题最有效的方法就是开发出真正有效的防攻击的水印技术。另外任何研制出高效水印技术的公司也应积极去寻找除去自己水印的方法并把它一同申请专利, 令盗版者无法利用这种消除水印的技术。

(2) 另一个重要问题就是水印探测器在系统中的位置。目前有两种主要的方案, 分别是将其置于 MPEG 解码器内部和置于 DVD 驱动器中, 采用哪种方案需要根据具体的应用环境和硬件条件来决定。

(3) DVD 播放设备应该具有对视频图像进行几何变换的能力, 例如将幅型比从 4:3 转换至 16:9。数字水印应该对这种变形有较强的抗干扰性。而且非法盗版者为了避开水印鉴别, 将会采取更为剧烈的尺寸变化或剪切等几何变换, 试图去掉视频图像中的水印。水印必须能经受住这类变换带来的影响, 这在不能使用帧缓存器的情况下显得特别困难。

(4) 视频压缩是去掉视觉上的不敏感信息, 这使得在图像高频部分嵌入水印十分困难。另一方面, 如果水印被放置于视觉敏感部位时, 原视频信息又难以进行有效的压缩。在 DVD 领域, 要求水印既可在压缩数据流中读出, 也可在重建的视频信息中读取。这使得数字水印必须能够经受 MPEG 的量化操作。而根据拷贝控制的需要, 压缩数据流中的水印应是可改变的 (包括增加或消除水印), 并且这种改变还不能影响码流的速率和 I 帧所处的位置。

(5) 水印鉴别是一种二进制操作, 因此有一个错误鉴别率的问题。如错误率太高, 将严重影响版权保护效果和合法用户权益。因此误码率应控制在 $1/10^{11} \sim 1/10^{12}$ 之间。

下面介绍一种将水印技术应用于 DVD 防复制控制的应用模型^[159], 下面给出了此模型的总体框架和各部分的详细分析。

9.5.3 一种典型的 DVD 复制控制系统方案

一种典型的复制控制系统基于 3 个基本概念, 它们是实现该系统的基础具体介绍如下。

(1) 数字水印: 包括嵌入在非压缩域的像素域水印和嵌入压缩域的 DC 系数水印。像素域水印能很好的抵抗模/数转换, 即使播放设备不与 DVD CSS 或者其他任何复制控制系统相容, 它也能正常工作; 而压缩域水印对于数字压缩有很好的表现, 它能保证重放控制在压缩域最大限度的起作用。

(2) 物理标志 P : 它是存储媒介的唯一电子标识。数字媒体必须包含有能鉴别此盘片是否为正规厂家生产的正版内容的标识, 为了这个目的, 每一个 DVD 盘片必须含有一个表明厂商的物理标志 P , 我们还引入了一个表明生产序列号的 K 的组合, 它和 P 组合起来共同构成此盘片唯一的标识。

(3) 授权标志 T : 含有加密了的 CGMS 信息, 表明此 DVD 的复制权限。它配合水

印一起起到控制复制的作用。

1. 复制控制和物理标志 P

DVD 播放器或驱动器最基本的一个功能就是应该测试内容的版权, 并拒绝播放盗版的内容。为了达到这个目的, 光盘必须包含一个可区别版权的标志。一种改进方案是: 根据厂商提供的种子 U , 利用单向方程产生 P , $P=F(U)$ 。当物理标志的比特内容被加入到盘片中后, 就不能从盘片外部来读取或恢复它。然后, 再根据 P 产生嵌入水印的密钥, 这样就将物理标志 P 与嵌入到内容中的水印联系到了一起, $W=\text{Insert}W(P)$ 。在 P 的产生过程中, 使用了 $y=F(x)$ 这个单向加密算法。这个单向方程必须满足的要求是若想逆向从 x 推出 y , 在计算量上是不现实的。注意到单向方程的计算只能在驱动器和记录器中进行, 一般使用简单的硬件解决方法。可以使用一个简单的电路, 只需要有很少的门数目, 它或者可以被递归的激发以增加强度, 或仅仅只是增加穷举搜索攻击的工作量, 目的就是使攻击者因为过大的工作量而放弃攻击。从安全性的观点来看, 这个单向方程的算法本身应该保密。

对于小规模盗版者来说, 当他想要让厂家刻录一定数量的盘片时, 他将遇到的一个困难就是找到物理标志 P 的比特值。而且, 他还需要提供给厂家产生 P 的种子 U , 而不仅仅提供物理标志 P 。从密码学的角度来说, 从物理标志 P 计算种子是不现实的, 因为使用的是单向方程。通过选择合适的单向方程, 可以减少以后对种子和物理标志的误用。在实验中, 可利用软件的方法来模拟物理标志 P 的功能, 并选择 MD5 单向哈希算法作为产生 P 的方程, $P=F(U)$, 再将 P 以密文的方式加入到视频文件头数据中的 INFO 块的 INAM 字段中, U 为厂家提供的唯一的种子。验证时根据厂家提供的种子表, 计算出所有的 $P'=F(U)$ 。若存在一个 $P'=P$, 则认为此视频文件为正版内容, 允许播放与传输; 否则, 判定此视频文件为盗版, 给出警告, 停止播放与传输。为了给每一个视频文件一个唯一的标志, 还可使用序列号 K 。序列号 K 的作用是赋给每一个视频文件一个唯一的标识。同样利用 MD5 算法计算出 K' , 然后将 P' 和 K' 结合起来, 一起作为密钥控制产生的鲁棒性水印序列。这样每一个视频文件都嵌入一个对应于本文件的唯一的水印序列, 而不会产生相同的厂家的视频文件都嵌入同一个水印序列的情况, 提高了水印的抗攻击性。在系统中, 将 K' 嵌入到视频文件头中的 INFO 块的 IKEY 数据中。该方案的另一概念是引进对“一次复制”权限的改变控制。这是一种可以传递的复制控制系统, 如图 9.14 所示, 它允许普通的消费者从原始盘中复制任何数量的第一代复制, 而却将所有的第一代复制盘的复制权限改为“不许复制”, 使得不能从第一代复制盘再进行第二代复制。记录设备必须能检测到“一次复制”状态并将第二代复制盘的状态调整为“不许复制”。可以用一个加密的授权标志 T 来实现这个功能。 T 可以存储在盘片中, 并合法的被复制或传输(通过卫星或电缆)。关于授权标志 T , 在下面详细讨论。

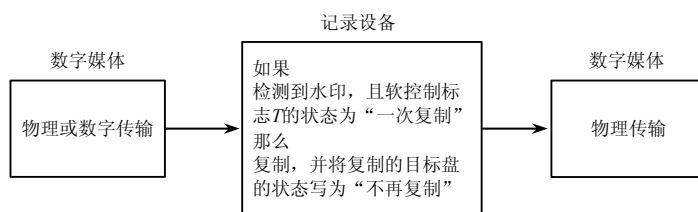


图 9.14 复制控制的基本思想

2. 授权标志 T : 一个基于密码学的安全的 CGMS 系统

在 DHS (Data Hiding Subgroup) 提出的 CFP (Call For Proposals) 系统中, 一个“一次复制”状态含有以下的意义: 通过加入第二个水印或者调整原先的水印来表明已经进行过一代复制, 并禁止下一代的复制。这种方法的缺点是第二个水印很容易受到攻击: 消费者和攻击者可以通过简单的比较“一次复制”和“不再复制”的两个图像序列来得到关于第二个水印的足够充分的信息, 并使攻击者可以编写代码将含有“不再复制”性质的盘片写回为“一次复制”状态。

利用基于授权标志的系统, 就能很好的克服上述这个缺点。通过讨论含有“一次复制”标志的 DVD-ROM 盘片来详细的探讨标志的作用。当允许从盘片上复制内容时, 驱动器必须输出一个授权标志。这个标志必须允许一个播放器播放复制的内容, 然后一个记录器复制内容。但在这些传输结束后, 不再允许下一代的复制。当每一次通过回放和记录设备之间时, 标志都改变自己的状态。这种状态的改变是通过加密的, 且是不可逆的, 它的作用是减少下一代可以复制和回放的次数。标志 T 包含 SCMS/CGMS 信息, 加在视频流头部, 表明还允许几代的复制, 也就是还可以做几次重放和记录。有四种状态: “一次复制”、“不再复制”、“不许复制”和“自由复制”。对于“不许复制”状态, T 表明只允许重放。对于“一次复制”状态, 盘片中的 T 有三节: 重放、记录、重放。当从播放器传输到记录器, 这个标志将被剪去一节。标志作为一个加密的计数器, 只能减少不能增加。计数器值的减少是通过将比特值通过一个单向方程来实现的。这表明每一次经过记录或重放操作, 视频流中的标志 T 就将被 $T' = F(T)$ 来代替, 这里 $F(\cdot)$ 是一个单向方程。没有播放器或者记录器能够在不改变 T 的状态的条件下进行操作。

通过比较不同的方案, 建议将标志 T 存储在 DVD-ROM/Video, DVD-R 和 DVD-RAM 的导入区数据中。在 DVD-ROM/VIDEO 中, 在导入区的数据含有与复制控制有关的内容。若使用了加密措施, 则在其含有的 16 个扇区的第 3 个扇区中含有 CSS (内容加扰系统) 的密钥表。若不考虑加密的话, 我们可以将标志放在第 4 个扇区中。DVD-R 也含有一个相似的结构。所以可以按照与 DVD-ROM/VIDEO 一样的方式来存储授权标志。而 DVD-RAM 有一个稍微不同的结构。它的导入区含有三个区: 一个不可重写区域, 一个镜像区域以及一个可重写区域。因为在不可重写区域中, 含有一个结构与在 ROM 和 R 中的控制数据一样的控制数据块, 所以, 这个区不能被用于携带授权标志。而镜像区也不能用。所以, 位于引入区中 30000h-31000h 的可重写区域是最适合的。对 RAM 来说, 第 0 个逻辑扇区对应着物理扇区的 31000h (对于 ROM 和 R 来说, 是 30000h)。在一个 DVD-RAM 盘片的引入区的可重写区域中, 我们发现有几个区: guard track 区、disc test 区、drive test 区、又一个 guard test 区、Disc Identification 区、Defect Management 区域。可以使用 disc identification 区来加入标志。这种方案保证了导入区的数据可以在驱动器里被正确的处理。并且标志可以获得足够多的存储空间, 使得也可以加入多重标志。

对于通过卫星或电缆的 MPEG 传输, 标志可以位于 MPEG 的用户数据区。对于模拟视频, 标志可以作为一个高频嵌入部件被传输, 原理与脆弱水印相似。普通的 VCR 可能会破坏这个信号, 使它丢失, 但这也使得这个视频内容自动成为“不再复制”状态。

在实验系统中, 可以用软件的方法来模拟授权标志, 所以可将其叫做软控制标志 T 。并将 T 和水印结合起来标示视频内容的复制权限。当检测到水印的时候, 根据 64 位的软控制标志 T 来判定此视频内容是哪种复制状态; 而当没有检测到水印的时候, 判定此视频文件为“不许复制”状态。有些文献与这里判定“自由复制”和“不许复制”的

策略正好相反：将不含有水印的内容判定为“自由复制”。但这种方法在攻击者对加入水印的内容进行恶意攻击，以使水印内容不可识别时不具有鲁棒性。这里，将 T 嵌入到视频文件头中的 INFO 块的 ICOP 数据中，且满足 $P=F(U)$ 、 $T=F(F(U))$ 。因为 T 是由 P 和 U 决定的，因此一个特定文件的标志不能被用于另一个文件。

3. 水印方案

在复制控制系统框架中，针对不同的压缩域应采用不同的水印方案，如图 9.15 所示。在非压缩域中（AVI 文件），嵌入像素域水印；在压缩域中（MPEG 文件），在 I 帧中的直流系数中嵌入水印。像素域水印只在 MPEG 解码器和编码器之间有效，它传递版权信息；DC 系数水印在编码器和解码器之间有效。

当视频流处于非压缩状态的时候，需要使用像素域水印的方法来判断文件的复制状态；而当文件被进行了压缩编码之后，即使像素域水印遭到了破坏，也可以使用压缩域的 DC 系数水印方案来判断是否存在水印。

简单的来说，像素域水印是通过改变视频序列帧像素的颜色值来加水印，以使其能在模拟信号域被检测出来。像素域水印方法是将水印加入一个视频序列的所有帧内。而在检测时，并不需要检查视频流中的所有帧，检查多少帧将依赖于计算机的速度。而压缩域水印是通过改变 MPEG 码流中的 I 帧内的 DC 系数值的色度分量值来加入水印的。我们将水印加入压缩后的视频流中的所有 I 帧中，同非压缩水印一样，在检测的时候，提取出多少帧的水印也是依赖于计算机的速度和应用环境。另外，在非压缩域，因为在每一帧中嵌入的鲁棒性水印都是一样的，所以不能很好地抵抗统计攻击中的共谋攻击。因为在所有帧中嵌入相同的水印，攻击者可以从单个的帧中估计出水印，并在不同的场景中求平均以取得较好的精确度，接着从每帧中减去估计的水印。因此，应该使用双重水印机制：在每一帧中除了嵌入鲁棒性水印外，还嵌入与内容相关的脆弱水印，一旦检测不出脆弱水印，就可以判断原视频内容遭到了攻击。这样就可以有效地抵抗鲁棒性水印不能抵抗的统计攻击。

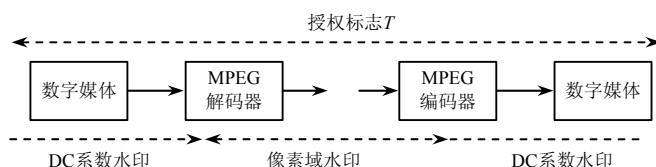


图 9.15 水印方案

4. 水印检测器的位置

若将水印检测器置于 MPEG 解码器内部可使这两部分共享一些重要资源，如码表和存储器等。然而，这样做的弊端在于没法防止 DVD 驱动器读出盗版光盘，而且目前大量 MPEG-2 解码器并不具备检测水印的能力，所以在一段时间内仍然无法限制盗版光盘地流通。如果将水印检测器置于 DVD 驱动器中将更加安全。因为 DVD 驱动器不仅能判断盘片的类型（ROM 或 RAM），而且其中的回放控制和记录控制会直接读出光盘上受水印保护的信息，保证只有正版的光盘才能被正常播放。

但是，将水印检测器置入 DVD 驱动器要考虑成本。为了将改造费用降至最低，驱动器制造商认为水印检测器必须用驱动器中未充分利用的半导体器件，这表明用于 DVD 系统的水印检测装置不能太复杂，其实现电路在 3 万门左右。这意味着检测器为了实时处理视频的需要将不能利用帧缓存器件，这就要求水印算法具有很好的性能和速度。

5. 总体框架

综上所述，可以给出如图 9.16 所示的总体框架。

9.6 广播监控

广播电视是人们最常见、最广泛的视频形式，广播电视中会出现电视节目盗播、广告不能按约定的时间、数量播放等各种问题，对电视节目播出过程进行实时的不间断地监控，是保证节目准确播出的必要手段，而这些问题单纯靠人工来监控不但费时、费力，效果也不理想。因此，人们开始试着将数字水印运用到视频广播监控上来，期望能够通过数字视频水印实现监控的功能。下面首先介绍数字广播系统和视频广播的独特之处，然后介绍数字视频广播监控问题，最后介绍基于数字水印的视频广播监控典型方案^[160]。

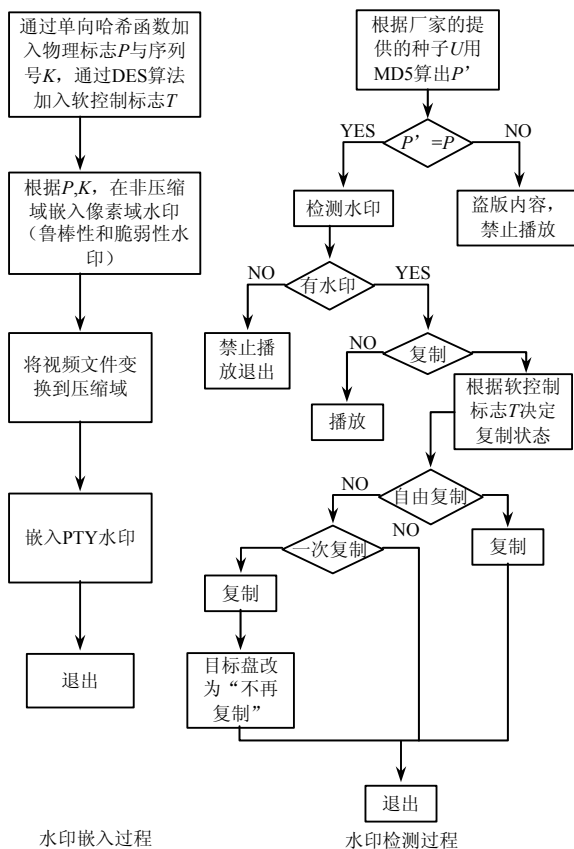


图 9.16 基于水印的 DVD 复制控制系统总体框架

9.6.1 数字视频广播

1. 数字广播系统

随着电视的数字化，极大的拓展了广播影视节目的传播，使广播电视节目能够更方便更快捷地传播，给广播影视节目的制作、传播和消费带来了全新的形态。对电视节目播出过程进行实时地不间断地监控，是保证节目准确播出的必要手段。电视媒体监测服务（电视新闻监测、电视广告监测、监播）已逐渐成了必不可少的信息服务。目前大部

分电视监测方法,在功能实现方面大多使用人工采集,采用的是用长时间的录像机将每天的节目录制下来,等到播出结束后,回放录制的磁带,来判断节目播出是否正常,在准确性、完整性、及时性各方面都存在弊端。

数字广播系统由信源播出系统、网络传输系统、条件接收系统、用户管理系统、网络管理系统及节目调度系统等组成。信源播出系统包括:证券播出系统、网站播发系统、文件播发系统、远程教育播出系统、媒体播出系统、视频播出系统、卫星电视接收、外交互的视频点播系统(NVOD)及INTERNET接入系统等组成。它们能够根据需要灵活组合,一台播控服务器可以播出多个不同种类信源。考虑数字广播系统平台发展需要,可从证券播控系统、远程教育播出系统、多媒体播出系统、文件及网站播出系统开始,逐步开展其他信源播出系统。用户管理系统完成功能包括:用户授权及管理、节目管理、资源管理、计费、账户管理等,同时实现了用户管理与条件接收系统的集成,完成节目的加密及用户确认等功能。节目调度系统完成节目的调度及控制,生成节目菜单并通知数字广播系统的DVB网关,生成PSI(SI)信息表。节目调度系统用图表方式直观显示节目播出计划及安排,操作人员在一的授权下可以随时修改节目播出计划,插播广告等。节目调度系统同时还具备一定的节目预览功能。网络管理系统完成网络的管理、节目的监控等功能,它包括:故障管理、配置管理、性能管理、安全管理等功能。网络传输系统包括数据广播网关、接口转换器、QAM调制器、复用器等设备,完成DVB数据形成、传输及调制等功能。上述各系统相互关联、相互结合构成一个完整数字广播系统平台,其相互关系如图9.17所示。

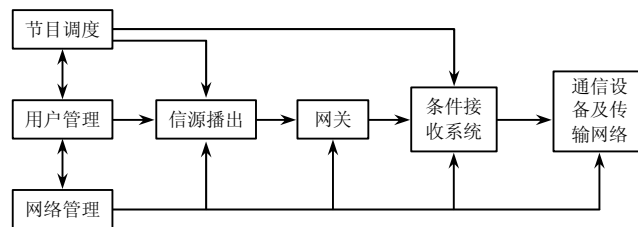


图 9.17 数字广播系统

2. 视频广播的独特之处

视频广播是从**数字声音广播**(Digital Audio Broadcast, DAB)的基础上发展而来的,与声音广播不同的是,视频广播不再是单纯声音广播,而是一种能同时传送多套声音节目、数据业务和活动图像节目的广播。它充分利用了数字音频广播技术优势,在功能上将传输单一的音频信号扩展为可传输数据文字、图形、电视等多种载体信息的信号。**数字多媒体广播**(Digital Multimedia Broadcast, DMB)是从数字声音广播 DAB 的基础上发展而来的,与 DAB 广播不同的是,DMB 广播不再是单纯声音广播,而是一种能同时传送多套声音节目、数据业务和活动图像节目的广播,它充分利用了 DAB 数字音频广播技术优势,在功能上将传输单一的音频信号扩展为可传输数据文字、图形、电视等多种载体信息的信号。

9.6.2 视频广播监控问题

视频广播监控系统是以计算机为核心,结合先进的多媒体技术、网络通信技术、数字图像压缩技术的一种远程视频监控系統。远程视频监控系統能将监控现场的监控信息

通过计算机网络传送到网络中的其他计算机上，并与信息管理系统一起达到远程监控的目的。广播电视中常会需要监视和校正数字电视（DTV）中 A / V 延时、视频传输中的检测控制、音视频的数字版权管理与广播监控。

1. A / V 延迟

A / V 延迟是自数字传输出现以来一直困扰着数字视频专业人员的问题。A / V 延迟的典型症状是节目的声音和画面不同步，其原因是在切换和演播室同步时数字视频处理发生画面延迟。例如，观众先听到新闻播音员的声音，然后才看到嘴形的相应变化。如果 A / V 延迟严重，便可能造成可怕的同步误差，影响服务视频质量，影响观众的欣赏效果，导致观众不满。虽然现在采用的“隶属于”视频同步器的音频同步器有助于维持固定的 A / V 定时关系，可以校正由播出机房的视频帧同步器引入的延时，但它对于在播出机房视频帧同步器的上游节目里的 A / V 延时变化引入的可变 A / V 延时误差，就无能为力。视频节目链路中串接在一起的设备所产生的累积误差，广泛的视频处理和运行不当的时间码，所有这一切都可能在音频和视频信号的相对定时上产生突然的漂移或逐渐的变化。

2. 视频传输中的检测控制

视频常在受限带宽的有误信道上进行传输，有误差信道又具有大量的随机误码和突发误码问题，而高压缩码流对信道误码非常敏感。为了保证高质量的视频通信传输，目前的信源信道编码分离的常用整体解决方案中，通常在视频解码端采用错误隐藏技术来改善错误发生时的图像重建。传统的基于语法检测的错误检测方案尽管简单易行，但是在检测量化后的 DCT（Q-DCT）系数的错误上存在重大缺陷，而视频传输中 Q-DCT 系数信息出错的概率又比较大，导致未被检测到的错误的宏块无法应用错误隐藏技术，结果重建图像的质量很差，再加上运动补偿技术的采用，帧内的错误将扩散到多帧重建图像中去，整个重建序列的质量将令人无法接受。

3. 音视频的数字版权管理与广播监控

数字水印技术可应用于开放环境下的数字广播影视的数字版权管理与广播监控。DS7510 广播电视网络综合管理系统是对空间及有线多套节目的质量和内容进行集中监测和综合管理的系统。主要进行单节点监测与控制，包括基本硬件测试平台、本地监控软件及数据库管理系统，可监测电视网络各节点的射频指标、调制质量、音视频指标，并采集记录播出的音视频内容，通过传输网络将传送数据汇总到网络监控和维护中心，以达到对整个有线运行网络维护和监测管理的目的。通过音频分析模块对音频信号载波质量，是否停播进行分析和判断，故障报警并上传给上位计算机，可对射频信号进行音视频解调并输出。利用视频分析模块 DS8611a 频谱分析仪，可测量范围为 9KHz-1GHz，输入信号范围为 10~110dBuV，测量精度为 $\pm 1.5\text{dBuV}$ ，具有频谱分析功能，可点频或扫频，用于监测非系统设置的非法频点，将测量结果上传给计算机。并具有声光报警功能，系统分析与控制管理中心对测量结果进行显示、存储、信号分析，判断停播劣播现象并报警输出，所选频道的运行情况以数据、曲线或棒状图方式实时显示。还可进行异态数据特殊警示历史记录查询，查询一年内各频道异态及报警情况，最近两月运行记录，以运行曲线方式显示。设操作员级、管理员级，赋予不同的操作权限，分别设密码保护，对监测时间、录像时间、异态判别、报警上下限阈值可随时设定和修改。异态、报警均有存储记录，可保存一年，运行数据保存一个月，生成数据库，并可查询历史记

录,生成报表、打印输出,通过故障输出语音指示灯报警,异态数据在明显位置给予特殊警示。

9.6.3 基于数字水印的视频广播监控

Nextamp 数字电视水印系统是法国 Nextamp 公司提供的数字电视监播解决方案。它采用数字水印技术实现对数字电视内容的版权保护和权益管理。电视节目在制作后在数字视频流或视频文件中插入一个 22~64bit 长的不可见的水印标识,加入水印标识可以被定义具有识别信息的水印标识,可以是一虚拟条形码,包括内容标识符、日期和时间号或加入时戳的台标图像等。水印具有隐藏性,嵌入到 MPEG-2 压缩或未被压缩的视频内以后水印标识是观众视觉上不可见的,同时水印标识具有不可擦除性,能够抵抗视频格式变换、编辑,标识插入和 MPEG 压缩与解压缩等操作,对于视频内容不会有任何影响,就是说水印标识具有自适应性,根据每幅图像的形状来对水印标识的不可见性和鲁棒性进行优化,这一不可见的标识始终能被读取,都可以找到它的源头,从而大大降低内容被盗用及盗版的可能性,解决内容提供商关注内容的安全性问题。该技术可以防止数字电视内容的非法使用,实现内容追踪,抗击非法视频攻击。系统流程如图 9.18 所示。

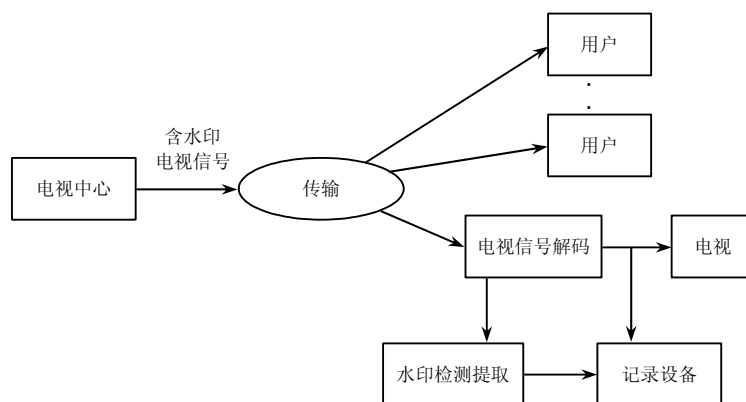


图 9.18 利用数字水印技术的广播监控系统

该系统通过识别水印能准确证明特定形式和长度的广告已实际播放,验证广告节目的完整性以及实际播放时段与付费时段的匹配性,提供在广播电视网络中播放广告的证据,确保播放广告的收入。对电视节目内容广播权进行全面验证,确保当前自动播放的节目与实际广播的视频节目原样一致,并对权利和元数据进行更新。对第三方频道全部或部分节目进行连续监测,保护体育和新闻等高品质节目内容的独家权利,统计节目内容被第三方频道引用或重新使用情况,持续追踪和检测已做标明的节目,及时记录节目内容运行状况。

广电节目审核机关可以在通过审核的所有视、音频加入包含各级播出、内容制作商的信息与审核标识水印,广电管理机构通过各级监测点采集所有有线无线电视广播、Internet 上需要监管视、音频信号,然后通过水印识别系统检测水印印章来自动监控所有视频传输内容,及时发现非法音像。不仅有效防止对审核过的节目内容篡改,而且可以加强对非法流通、传播的控制与管理。解决对非法视频信号干扰误播问题。数字水印技术不同于加密技术,如果在开放的免费频道节目发送、接收端采用加解密加密技术,普

通用户需要专用解密设备, 由于成本、技术原因, 必然失去这部分观众 (大量的小团体、个体接收用户), 从而失去开放频道的意义。

通过在数字视频内容中加入数字水印, 可以在防止非法数字电视内容的使用, 实现内容追踪, 抗击非法视频攻击, 但是不影响普通用户接收, 对于普通用户可以忽略数字水印的存在, 直接接收视频信号。例如广告商出资向电视台购买广告时段, 最关心的一个问题就是他的广告有没有如期如量播出。目前采用的是一种技术含量较低的方法: 派专人监视广播并记录其所见的内容。这种方法的不足之处在于效率不高且容易出错。若用自动监测系统取代监视人员, 可起到事半功倍的作用, 这种自动监测系统利用视频水印技术即可实现, 只要在广告片中嵌入自己的标识, 然后在接收端对电视信号进行检测, 检测到自己的标识水印, 则能证明该广告片被播出, 并可控制相应设备进行相关信息的记录, 从而实现智能化的广告监播。

9.7 其他应用

除了上述几个主要应用场合外, 信息隐藏技术还在许多方面得到应用, 下面列举几种其他主要应用场合。

9.7.1 印刷防伪

目前, 各个企业纷纷加大在品牌创立方面的投资。在不断扩大企业和产品的知名度上不遗余力。但是在名牌产品创立的同时, 各种各样的假冒伪劣产品也以惊人的速度和数量充斥市场。当前假冒伪劣商品的交易额约占世界贸易总额的 5%~6%, 每年高达 1500~1800 亿美元, 给名牌企业和名牌商品带来极大的破坏, 使企业蒙受巨大的损失。在我国, 假冒伪劣产品每年造成 2000 亿人民币以上的经济损失。因此, 我们必须加大防伪力度, 严防假冒伪劣产品对企业和消费者的伤害。

防伪技术已成为一个崭新的并在不断发展变化的技术, 其一般可分为物理防伪技术、化学防伪油墨技术、生物防伪技术、材料防伪技术、计算机网络防伪技术、防伪印刷技术、防伪包装技术等。我们平时接触的图像、文字大部分最终要通过印刷输出, 如商标、产品包装、书刊、证书、证件、邮票、磁卡、出版物、影音制品封面、货币、有价票券、入场券、单据、广告、挂历台历、名片、护照等, 这就使得防伪印刷技术显得非常重要。目前, 防伪印刷技术在我国是使用最普遍、识别最简单、成本最低的防伪手段。

从印刷技术的角度来看, 印刷防伪技术包括雕刻制版、用计算机设计版纹、凹版印刷、彩虹印刷、花纹对接、双面对印技术、多色接线印刷、多色叠印、缩微印刷技术、折光潜影、隐形图像和图像混扰印刷等。虽然人们在彩色印刷上研发了不少防伪方法如激光防伪、专用水印纸防伪、电话电码防伪、电脑网络防伪、特殊加网防伪、分期解密防伪、隐形标识防伪、隐形图像防伪等, 但是这些防伪方法只是在一定范围内起到了一定程度的防伪作用, 而产品的成本却大幅度增加, 也就是说名牌产品企业每年要为防止假冒花费很多资源, 所以人们更加渴望一种成本低廉的防伪方法。在这种需求下, 数字水印技术就可以很好地解决这些问题, 既能达到很好的防伪效果, 又能有效地降低成本。以商标防伪印刷系统为例说明: 首先在计算机中对商标进行设计, 然后在设计好的图像中加入唯一可以识别的水印信息, 最后经过排版印刷制作商标。当发现假冒产品时, 可以利用专用检测器对商标进行检测, 正版产品的商标中含有可以证明版权的水印

信息，而假冒商品中则没有。

与印刷品防伪稍有差别的是证件防伪。印刷品的防伪常常要求水印系统对一次扫描鲁棒，对二次扫描脆弱印刷品包括印刷的票据、书籍等。由于数字水印技术可以它们提供不可见的标志，从而大大增加了伪造的难度。证件的防伪则要求水印系统具有足够的鲁棒性。证件的防伪已有较为成熟的应用实例，如德国的汽车驾照就采用了水印技术。

9.7.2 软件保护

随着软件产业的不断发展，软件盗版、代码窃取以及篡改等问题同益严重，软件产品的版权保护受到越来越多地重视。目前，来自恶意主机对知识产权的攻击方式主要有三种：**软件盗版**（Software Piracy）、**逆向工程**（Reverse Engineering）、**代码篡改**（Code Tamper），一般来说，如果软件能被专利、版权、契约、贸易安全法等合法形式保护，那么可以阻止对软件的攻击。但很遗憾，发现破坏软件版权的行为十分困难，即使发现了违法破坏行为，通过法律手段来寻求补偿也是非常昂贵甚至是不可能的。所以，用技术手段来保护软件开发商的版权是非常必要而且具有很大应用价值。针对上述三种攻击形式，对应的防御措施为**软件水印技术**（Software Watermarking）、**代码模糊技术**（Code Obfuscation）、**软件防篡改技术**（Software Tamper-proofing）。

1. 特性

软件水印技术就是在程序 P 中嵌入水印信息 W 生成目标程序 P' ，其中 W 应该具有以下特征。

(1) 在目标程序 P' 遭受各种代码转换攻击之后，仍能将 W 可靠的查找和识别，并正确提取，即 W 在 P' 遭受攻击后具有可恢复性；

(2) P 和 P' 的静态特性应该无差异，功能应该没有明显的差异，即 W 需要具有隐蔽性和正确性；

(3) W 需要具有较高的数据率；

(4) W 应该具有某种数学特性，用来证明 W 在 P' 中的存在是人为的。

2. 分类

目前软件水印很多分类标准，一般可以按照以下几种方式进行划分。

(1) 根据水印加载的位置进行分类

根据水印被加载的位置，软件水印可以分为**代码水印**（Code Watermark）和**数据水印**（Data Watermark）两种。前者隐藏在可执行代码的指令部分，后者隐藏在数据中。

(2) 根据水印的保护功能分类

根据水印的保护功能分类可以分为如下四类。

1) **身份标识水印**（Authorship Mark）：通过在软件中嵌入作者的身份信息为其提供知识产权保护，这种水印用来标识作者的身份，一般隐藏于软件中，不但可以标识单一作者身份，也可以标识多个拥有者。

2) **验证水印**（Validation Mark）：通过生成文档的密码摘要来证明软件自发布之日起未被篡改过，如 Sun 公司在 2001 年开始支持对 Java 小程序进行数字签名，该种用来验证软件的所有信息，包括源代码、程序可执行文件、程序安装文件和程序资源文件（图像、图标或其他信息）等信息，仍然与软件发布时一样，而没有经过篡改，以验证发布者的真实性。

3) **许可证水印 (Licensing Mark)**: 主要用于表明软件规定的使用方式, 以配合验证软件是否被误用或盗用, 从而验证使用者的合法性。通过含有一个解密密钥, 处于许可控制下的软件被加密, 当许可水印被破坏时解密密钥随之失效, 以此来控制软件的使用。

4) **指纹水印 (Fingerprinting Mark)**: 通常是在软件中嵌入购买者信息和软件的序列号, 以防止非法复制或用于收集关于分发渠道的统计信息。这种水印必须在一定范围内包含一个独一无二的身份信息, 用来标识购买者 (用户) 的身份。

针对不同的应用, 上述各种水印在鲁棒性和可见性方面也有不同的要求。和数字水印领域类似, 身份标识水印和指纹水印要求是鲁棒的, 而验证和许可水印有脆弱性需求。

(3) 根据水印被加载时刻分类

根据水印被加载的时刻, 软件水印技术可以分为**静态软件水印技术 (Static Software Watermarking)**和**动态软件水印技术 (Dynamic Software Watermarking)**。

9.7.3 媒体桥

媒体桥 (Media Bridge) 技术开拓了访问 Internet 的一条新途径, 通过在杂志广告、产品包装、目录甚至各类票据中隐藏不可见的数字水印。用户只要将这些传统媒体放在**网络摄像机 (Web Camera)** 前, 媒体桥技术就可以直接将用户带到与印刷图像内容相关联的网络站点, 省去用户敲击键盘和单击鼠标的过程。例如将登有广告的杂志置于网络摄像机前, 媒体桥技术便会主即在计算机上显示出广告公司的主页和广告中产品的相关信息, 免去用浏览器在网上搜索的过程。这种方式可以使出版商、广告商和图像应用者增加产品附加值。

9.8 本章小结

本章主要介绍了信息隐藏技术的几个最重要的应用领域: 知识产权保护、军事保密通信、交易跟踪、真伪鉴别、复制控制、广播监控等。在具体应用中, 还有许多技术问题有待解决。随着信息隐藏技术和理论的进一步深入研究, 其应用领域必将越来越广。



习题

1. 请举例说明如何使用信息隐藏技术进行知识产权保护。
2. 试编写一个使用信息隐藏技术进行军事保密通信的 Matlab 程序。
3. 使用何种信息隐藏技术才能进行交易跟踪, 需要注意哪些问题?
4. 真伪鉴别应用场合对信息隐藏技术有哪些特殊要求?
5. 使用信息隐藏技术实现复制控制还有哪些技术问题需要解决?
6. 设想一个信息隐藏技术的应用领域, 并给出系统框图。

参 考 文 献

1. 汪小帆, 戴跃伟, 茅耀斌. 信息隐藏技术-方法与应用[M]. 机械工业出版社, 2001.
2. Schneier B 著, 吴世忠, 祝世雄, 张文政等译. 应用密码学-协议、算法与 C 源程序[M]. 机械工业出版社, 2000.
3. Katzenbeisser S, Petitcolas F A P. Information Hiding Techniques for Steganography and Digital Watermarking[M]. Artech House, Inc., 2000.
4. 刘振华, 尹萍编著. 信息隐藏技术及其应用[M]. 科学出版社, 2002.
5. Cox I J, Miller M L, Bloom J A. Digital Watermarking[M]. Morgan Kaufmann, 2002.
6. Simmons G J. The History of Subliminal Channels[C]. Proceedings of the First International Workshop on Information Hiding, Lecture Notes in Computer Science, Vol. 1174, 1996, pp. 237~256.
7. 娄振华. 信息隐藏的安全性研究[D]. 解放军信息工程大学硕士学位论文. 2008.
8. Cox I J, Miller M L, Mckellips A L. Watermarking as Communications with Side Information [J]. Proceedings of the IEEE, 1999, 87(7): 1127~1141.
9. 林代茂, 胡嵐, 郭云彪, 周琳娜. 广义信息隐藏技术的机理与模型[J]. 北京邮电大学学报, 2005, 28(1): 1~5.
10. 李欣. 信息隐藏模型和若干问题研究[D]. 北京邮电大学博士学位论文, 2012.
11. 徐挺挺. 音频信息隐藏技术的研究与应用[D]. 南京邮电大学硕士学位论文, 2010.
12. Moulin P, O'sullivan J A. Information-theoretic Analysis of Information Hiding[J]. IEEE Transactions on Information Theory, 2003, 49(3): 563~593.
13. Cachin C. An Information-theoretic Model for Steganography [J]. Information and Computation, 2004, 192(1): 41~56.
14. 许冬艳. 基于中文句型变换的信息隐藏技术的研究与实现[D]. 西北大学硕士学位论文, 2009.
15. 李丽娟, 熊淑华. 基于文本的信息隐藏技术研究[J]. 现代电子技术, 2006.
16. Bender W, Gruhl D, Morimoto N, Lu A. Techniques for data hiding[J]. IBM System Journal, 1996, 35(3/4): 313~336.
17. Sharp T. An Implementation of Key-based Digital Signal Steganography[C]. Proceedings of the 4th International Workshop on Information Hiding, Pittsburgh, PA, USA, Lecture Notes in Computer Science, 2001, Vol. 2137, pp.13~26.
18. Kawaguchi E, Eason R O. Principle and Applications of BPCS Steganography[C]. Proceedings of SPIE Multimedia Systems and Applications, Boston, MA, USA, 1998, Vol. 3528, No.1, pp. 464~473.
19. Marvel L M, Boncelet C G, Retter C T. Reliable Blind Information Hiding for Images[C]. Proceedings of the 2nd International Workshop on Information Hiding, Portland, Oregon, USA, Lecture Notes in Computer Science, 1998, Vol. 1525, pp.48~61.
20. Wu D C, Tsai W H. A Steganographic Method for Images by Pixel Value Differencing[J]. Pattern Recognition Letters, 2003, 24(9~10): 1613~1626.
21. Zhang X P, Wang S Z. Vulnerability of Pixel-value Differencing Steganography to Histogram Analysis and Modification for Enhanced Security[J]. Pattern Recognition Letters, 2004, 25(3): 331~339.
22. Yu Y H, Chang C C and Hu Y C. Hiding Secret Data in Images via Predictive Coding[J]. Pattern Recognition, 2005, 38(5): 691~705.

23. Zhang X P and Wang S Z. Steganography Using Multiple-base Notational System and Human Vision Sensitivity[J]. IEEE Signal Processing Letters, 2005, 12(1): 67~70.
24. Koch E, Zhao J. Towards Robust and Hidden Image Copyright Labeling. Proceedings of IEEE Workshop on Nonlinear Signal and Image Processing, Neos Marmaras, Greece, 1995, pp.452~455.
25. Westfeld A . High Capacity Despite Better Steganalysis (F5-A Steganographic Algorithm). Proceedings of the 4th International Workshop on Information Hiding, Lecture Notes in Computer Science, Vol. 2137, 2001, pp. 289~302.
26. Sallee P. Model-Based Steganography[C]. Proceedings of International Workshop on Digital Watermarking, Berlin: Springer-Verlag, 2004, pp. 154~167.
27. Witten I H, Neal R M, Cleary J G. Arithmetic coding for data compression[J]. Communications of the ACM, 1987, 30(6): 520~540.
28. Chen B, Gregory W. Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding[J]. IEEE Transactions on Information Theory, 2001, 47(4): 1423~1443.
29. Abdulaziz N, Pang K K. Wavelet Transform and Channel Coding for Data Hiding in Video [C]. Proceedings of International Conference on Info-tech and Info-net, 2001, pp. 791~796.
30. Pazarci M, Dipcin V. Data Embedding in Scrambled Digital Video[C]. Proceedings of the Eighth IEEE International Symposium on Computers and Communication. 2003, pp.498~503.
31. Joumaa H, Davoine F. An ICA Based Algorithm for Video Watermarking[C]. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. II, 2005: 805~808.
32. Watson A B. DCT Quantization Matrices Visually Optimized for Individual Images[C]. SPIE Proceedings on Human Vision, Visual Processing, and Digital Display IV, Vol.1913, 1993, pp. 202~216.
33. 孙圣和, 陆哲明, 牛夏牧等著. 数字水印技术及应用. 科学出版社, 2004, pp.1~714.
34. Van Schyndel R G, Tirkel A Z, Mee N, Osborne C F. A Digital Watermark. Proceedings of IEEE International Conference on Image Processing, Austin, November 1994, Vol.2, 86~90.
35. Shih F Y, Wu S Y T. Combinational Image Watermarking in the Spatial and Frequency Domains. Pattern Recognition, 2003, 36(4): 969~975.
36. Alturki F , Mersereau R . Secure Fragile Digital Watermarking Technique for Image Authentication. International Conference on Image Processing, 2001, Vol. 3, pp.1031~1034.
37. Lie W N, Chang L C. Robust and High-Quality Time-Domain Audio Watermarking Subject to Psychoacoustic Masking. IEEE International Symposium on Circuits and Systems, 2001, Vol.2, pp.45~48.
38. 王秋生, 孙圣和. 一种在数字音频信号中嵌入水印的新算法. 声学学报, 2001, 26(5): 464~467.
39. Lacy J, Quackenbush S R, Reibman A R, Shur D, Snyder J H. On Combining Watermarking with Perceptual Coding. IEEE International Conference on Acoustics, Speech and Signal Processing, 1998, Vol.6, pp.3725~3728.
40. Qiao L , Nahrstedt K . Non-Invertible Watermarking Methods for MPEG Encoded Audio[C]. Proceedings of SPIE, 1999, Vol. 3675, pp.194~202.
41. Lu Z M, Ge Q M, Niu X M. Robust Adaptive Video Watermarking in the Spatial Domain[C]. Proceedings of the 5th International Symposium on Test and Measurement, Shenzhen, China, 2003, pp. 1875~1880.
42. Ge Q M, Lu Z M, Niu X M. Oblivious Video Watermarking Scheme with Adaptive Embedding Mechanism[C]. Proceedings of the Second International Conference on Machine Learning and Cybernetics, Xi'an, China, 2003, pp. 2876~2881.
43. Hartung F, Girod B. Digital Watermarking of MPEG-2 Coded Video in the Bitstream Domain[C]. Proceedings of IEEE International Conference on Acoustic, Speech, and Signal

- Processing, Munich, Germany, 1997, pp.2621~2624.
44. Simitopoulos D, Tsaftaris S A, Boulgouris N V, Strintzis M G. Compressed-domain Video Watermarking of MPEG Streams[C]. Proceedings of IEEE International Conference on Multimedia and Expo, Lausanne, Switzerland, 2002, pp. 569~572.
 45. Kutter M, Jordan F, Ebrahimi T. Proposal of a Watermarking Technique for Hiding Retrieving Data in Compressed and Decompressed Video[R]. Technical Report M2281, ISO/IEC Document, JTC1/SC29/WG11, Stockholm: MPEG-4 Meeting, 1997.
 46. Hartung F, Eisert P, Girod B. Digital Watermarking of MPEG-4 Facial Animation Parameters[J]. Computer & Graphics, 1998, 22(3): 425~435.
 47. Lu C S, Chen J R, Liao M H Y, Fan K C. Real-time MPEG-2 Video Watermarking in the VLC Domain[C]. Proceedings of the 16th International Conference on Pattern Recognition, Guebec, Canada, 2002, pp.552~555.
 48. Wagner N R. Fingerprinting. Proceedings of IEEE Symposium on Security and Privacy[C], Oakland, CA, 1983, pp.18~22.
 49. Blakley G R, Meadows C, Purdy G B. Fingerprinting Long Forgiving Messages[C]. Proceedings of Advances in Cryptology, Lecture Notes in Computer Science, 1986, Vol.218, pp.180~189.
 50. Boneh D, Shaw J. Collusion-Secure Fingerprinting for Digital Data[J]. IEEE Transactions on Information Theory, 1998, 44(5): 452~465.
 51. Chor B, Fiat A, Naor M. Tracing Traitors[C]. Proceedings of the 14th Annual International Cryptology Conference on Advances in Cryptology, Lecture Notes in Computer Science, 1994, Vol. 839, pp. 257~270.
 52. Pfitzmann B, Schunter M. Asymmetric Fingerprinting[C]. Proceedings of the 15th Annual International Conference on Theory and Application of Cryptographic Techniques, Lecture Notes in Computer Science, 1996, Vol.1070, pp.84~95.
 53. Pfitzmann B, Waidner M. Anonymous Fingerprinting[C]. Proceedings of the 16th Annual International Conference on Theory and Application of Cryptographic Techniques, Lecture Notes in Computer Science, Vol. 1233, 1997, pp.88~102.
 54. Wu M, Trappe W, Wang Z J, Liu K J R. Collusion-Resistant Fingerprinting for Multimedia [J]. IEEE Signal Processing Magazine, 2004, 21(2): 15~27.
 55. 李新伟. 抗共谋攻击的数字指纹技术研究[D]. 西安电子科技大学博士论文, 2011.
 56. Cox I J, Kilian J, Leighton F T, Shamoon T. Secure Spread Spectrum Watermarking for Multimedia[J]. IEEE Transactions on Image Processing, 1997, 6(12): 1673~1687.
 57. Kilian J, Leighton F T, Matheson L R, Shamoon T G, Tarjan R E, Zane F. Resistance of Digital Watermarks to Collusive Attacks[C]. Proceedings of IEEE International Symposium on Information Theory, Cambridge, MA, 1998, pp. 271~291.
 58. Ergn F, Kilian J, Kumar R. A Note on the Limits of Collusion-Resistant Watermarks[C]. Proceedings of the 17th International Conference on Theory and Application of Cryptographic Techniques, Lecture Notes in Computer Science, 1999, Vol. 1592, pp. 140~149.
 59. Zane F. Efficient Watermark Detection and Collusion Security[C]. Proceedings of the 4th International Conference on Financial Cryptography, Lecture Notes in Computer Science, 2001, Vol. 1962, pp.21~32.
 60. Domingo-Ferrer J. Anonymous Fingerprinting of Electronic Information with Automatic Identification of Redistributors[J]. Electronics Letters, 1998, 34(13): 1303~1304.
 61. Domingo-Ferrer J, Herrera-Joancomarti J. Efficient Smart-card Based Anonymous Fingerprinting [C]. The 3rd International Conference on Smart Card Research and Advanced Applications, Louvain La Neuve, Belgium, Lecture Notes in Computer Science, 1998, Vol.1820, 221~228.
 62. Domingo-Ferrer J. Anonymous Fingerprinting Based on Committed Oblivious

- Transfer[C]. Second International Workshop on Practice and Theory in Public Key Cryptography, Kamakura, Japan, Lecture Notes in Computer Science, 1999, Vol.1560, pp. 43~52.
63. Chung C, Choi S, Choi Y, Won D. Efficient Anonymous Fingerprinting of Electronic Information with Improved Automatic Identification of Redistributors[C]. Proceedings of the Third International Conference on Information Security and Cryptology, Lecture Notes in Computer Science, 2001, Vol. 2015, pp. 221~234.
64. Choi J G, Sakurai K, Park J H. An Efficient Fingerprinting Scheme with Proxy Signatures[C]. Proceedings of the Symposium on Cryptography and Information Security, 2003, pp. 1151~1156.
65. Pfitzmann B, Sadeghi A R. Coin-based Anonymous Fingerprinting[C]. Proceedings of the 17th International Conference on Theory and Application of Cryptographic Techniques, Lecture Notes in Computer Science, 1999, Vol. 1592, pp. 150~164.
66. Pfitzmann B, Sadeghi A R. Anonymous Fingerprinting with Direct Non-repudiation[C]. Proceedings of the 6th International Conference on the Theory and Application of Cryptology and Information Security: Advances in Cryptology, Lecture Notes in Computer Science, 2000, Vol. 1976, 401~414.
67. Kuribayashi M, Tanaka H. A New Anonymous Fingerprinting Scheme with High Enciphering Rate[C]. Proceedings of the Second International Conference on Cryptology in India: Progress in Cryptology, Lecture Notes in Computer Science, 2001, Vol. 2247, pp.30~39.
68. Okamoto T, Uchiyama S. An Efficient Public-key Cryptosystem[C]. Proceedings of the International Conference on the Theory and Application of Cryptographic Techniques, Lecture Notes in Computer Science, 1998, Vol. 1403, pp.308~318.
69. Camenisch J. Efficient Anonymous Fingerprinting with Group Signatures[C]. Proceedings of the 6th International Conference on the Theory and Application of Cryptology and Information Security: Advances in Cryptology, Lecture Notes in Computer Science, 2000, Vol. 1976, pp. 415~428.
70. Safavi-Naini R, Wang Y. Anonymous Traceability Schemes with Unconditional Security[C]. Proceedings of the First International Conference on Progress in Cryptology, Lecture Notes in Computer Science, 2000, Vol. 1977, pp. 250~261.
71. Chaum D. Blind Signatures for Untraceable Payments[C]. Advances in Cryptology: Proceedings of CRYPTO'82, 1983, pp.199~203.
72. 孙中伟, 冯登国. 一种基于同态公钥加密体制的匿名数字指纹方案[J]. 中山大学学报(自然科学版), 2004, 43(11增刊2): 109~112.
73. Fridrich J J, Goljan J, Du R. Invertible Authentication[C]. Proceedings of the SPIE, Security and Watermarking of Multimedia Content III, 2001, Vol. 4314, pp. 197~208.
74. Goljan M, Fridrich J J, Du R. Distortion-Free Data Embedding for Images[C]. Proceedings of the 4th Information Hiding Workshop, Pittsburgh, Pennsylvania, Lecture Notes in Computer Sciences, 2001, Vol. 2137, pp. 27~41.
75. Celik M U, Sharma G, Tekalp A M, Saber E. Reversible Data Hiding[C]. Proceedings of International Conference on Image Processing, 2002, Vol. II, pp. 157~160.
76. Tian J. Reversible Watermarking by Difference Expansion[C]. Proceedings of the Workshop on Multimedia and Security, 2002, pp.19~22.
77. Thodi M, Rodriguez J J. Reversible Watermarking by Prediction-Error Expansion[C]. Proceedings of the 6th IEEE Southwest Symposium on Image Analysis and Interpretation, Lake Tahoe, USA, 2004, pp. 21~25.
78. Ni Z, Shi Y Q, Ansari N, Su W. Reversible Data Hiding[C]. Proceedings of IEEE International Symposium on Circuits and Systems, 2003, Vol. II, pp. 912~915.
79. Yang B, Schmucker M, Funk W, Busch C, Sun S. Integer DCT-Based Reversible Watermarking for Images Using Companding Technique[C]. Proceedings of SPIE, Security, Steganography, and Watermarking of Multimedia Contents VI, Proceedings of the SPIE, 2004,

- Vol. 5306, pp. 405~415.
80. Plonka G, Tasche M. Invertible Integer DCT Algorithms[J]. Applied and Computational Harmonic Analysis, 2003, 15 (1): 70~88.
 81. Xuan G, Yang C, Zhen Y, Shi YQ, Ni Z. Reversible Data Hiding Using Integer Wavelet Transform and Companding Technique[C]. Proceedings of 2004 International Workshop on Digital Watermarking, Lecture Notes in Computer Science, 2005, Vol. 3304, pp. 115~124.
 82. Cohen A, Daubechies I, Feauveau J C. Biorthogonal Bases of Compactly Supported Wavelets[J]. Communications on Pure and Applied Mathematics, 1992, 45 (5): 485~560.
 83. Yang B, Schmucker M, Niu X, Busch C, Sun S. Reversible Image Watermarking by Histogram Modification for Integer DCT Coefficients[C]. Proceedings of the 6th IEEE Workshop on Multimedia Signal Processing, Siena, Italy, 2004, pp. 143~146.
 84. Fridrich J, Goljan M, Du R. Invertible Authentication Watermark for JPEG Images[C]. Proceedings of International Conference on Information Technology: Coding and Computing, Las Vegas, Nevada, USA, 2001, pp. 223~227.
 85. Hong W, Chen T S, Shiu C W. Lossless Steganography for AMBTC Compressed Images[C]. Proceedings of the 1st International Congress on Image and Signal Processing, 2008, pp. 13~17.
 86. Chen J, Hong W, Chen T S, Shiu C W. Steganography for BTC Compressed Images Using No Distortion Technique[J]. The Imaging Science Journal, 2010, 58 (4): 177~185.
 87. Li C H, Lu Z M, Su Y X. Reversible Data Hiding for BTC-Compressed Images Based on Bitplane Flipping and Histogram Shifting of Mean Tables[J]. Information Technology Journal, 2011, 10 (3): 1421~1426.
 88. Chang C C, Lin C C, Tseng C S, Tai W L. Reversible Hiding in DCT-based Compressed Images [J]. Information Sciences, 2007, 177 (13): 2768~2786.
 89. Xuan G, Shi Y Q, Ni Z, Chai P, Cui X, Tong X. Reversible Data Hiding for JPEG Images Based on Histogram Pairs [C]. Proceedings of the 4th International Conference on Image Analysis and Recognition, Montreal, Canada, 2007, pp.715~727.
 90. Fridrich J, Goljan M, Chen Q, Pathak V. Lossless Data Embedding with File Size Preservation [C]. Proceedings of SPIE, Security and Watermarking of Multimedia Contents VI, San Jose, California, 2004, pp.354~365.
 91. Wang K, Lu Z M, Hu Y. A High Capacity Lossless Data Hiding Scheme for JPEG Images[J]. Journal of Systems and Software, 2013, 86 (7): 1965~1975.
 92. 王永吉, 吴敬征, 曾海涛, 丁丽萍, 廖晓锋. 隐蔽信道研究. 软件学报, 2010, 21 (9): 2262~2288.
 93. 李恕海. 阙下信道与封闭阙下信道研究. 西安电子科技大学硕士学位论文, 2005.
 94. 董庆宽. 阙下信道技术研究. 西安电子科技大学博士学位论文, 2003.
 95. 宋虹. 匿名通信关键技术与通用体系结构研究. 中南大学博士学位论文, 2010.
 96. Westfeld A, Pfitzmann A. Attacks on Steganographic Systems—Breaking the Steganographic Utilities Ezstego, Jsteg, Steganos, and S-tools and Some Lessons Learned[C]. Proceedings of the 3rd Information Hiding Workshop, Dresden, Germany, Lecture Notes in Computer Sciences, 1999, Vol.1768, pp. 61~76.
 97. Fridrich J, Goljan M, Du R. Reliable Detection of LSB Steganography in Color and Grayscale Images [C]. Proceedings of 2001 ACM Workshop on Multimedia and Security: New Challenges, Ottawa, Ontario, Canada, 2001, pp. 27~30.
 98. Harmsen J, Pearlman W. Steganalysis of Additive Noise Modelable Information Hiding[C]. Proceedings of IS&T / SPIE Electronic Imaging: Security, Steganography, and Watermarking of Multimedia Contents V, San Jose, CA, USA, 2003, SPIE, Vol.5020, pp.131~142.
 99. Ker A D. Steganalysis of LSB Matching in Grayscale Images[J]. IEEE Signal Processing Letters, 2005, 12 (6): 441~444.

100. Pevny T, Bas P, Fridrich J. Steganalysis by Subtractive Pixel Adjacency Matrix[J]. IEEE Transactions on Information Security and Forensics, 2010, 5 (2): 215~224.
101. Provos N, Honeyman P. Detecting Steganographic Content on the Internet[C]. Proceedings of ISOC NDSS' 02, San Diego, CA, USA, 2002.
102. Lee K, Westfeld A, Lee S. Category Attack for LSB Steganalysis of Jpeg Images[C]. Proceedings of the 5th International Workshop on Digital Watermarking, Jeju Island, Korea, Lecture Notes in Computer Sciences, 2006, Vol. 4283, pp.35~48.
103. Fridrich J, Goljan M, Hoge D. New Methodology for Breaking Steganographic Techniques for Jpegs[C]. Proceedings of IS&T / SPIE Electronic Imaging: Security and Watermarking of Multimedia Contents V, San Jose, CA, USA, 2003, SPIE, Vol. 5020, pp.143~155.
104. Fridrich J, Goljan M, Hoge D. Steganalysis of Jpeg Images: Breaking the F5 Algorithm[C]. Proceedings of the 5th Information Hiding Workshop, Netherlands, Lecture Notes in Computer Sciences, 2002, Vol.2578, pp.310~323.
105. Fridrich J, Goljan M, Hoge D. Attacking the Outguess[C]. Proceedings of ACM Workshop on Multimedia and Security, France, 2002, pp.3~6.
106. Avcibas I, Memon N D, Sankur B. Steganalysis of Watermarking Techniques Using Image Quality Metrics[C]. Proceedings of the SPIE, Security and Watermarking of Multimedia Contents, New York, SPIE, 2001, pp.221~229.
107. Farid H. Detecting Hidden Messages Using Higher-order Statistical Models[C]. Proceedings of the IEEE International Conference on Image Processing, 2002, pp.905~908.
108. Fridrich J, Pevny T. Multi-class Blind Steganalysis for JPEG Images[C]. Proceedings of the SPIE Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents VIII, SPIE, 2006, pp.1~13.
109. Holotyak T, Fridrich J, Voloshynovskiy S. Blind Statistical Steganalysis of Additive Steganography Using Wavelet Higher Order Statistics[C]. Proceedings of the 9th IFIP TC-6/TC-11 Cone Communications and Multimedia Security, 2005, pp. 273~274.
110. Goljan M, Fridrich J, Holotyak T. New Blind Steganalysis and Its Implications[C]. Proceedings of the SPIE, Security, Steganography, and Watermarking of Multimedia Contents VI, SPIE, 2006, Vol. 6072, pp.1~13.
111. Xuan G, Shi Y Q, Gao J. Steganalysis Based on Multiple Features Formed by Statistical Moments of Wavelet Characteristic Functions[C]. Proceedings of the 7th International Workshop on Information Hiding, 2005, pp.262~277.
112. Shi Y Q, Xuan G, Zou K. Steganalysis Based on Moments of Characteristic Functions Using Wavelet Decomposition, Prediction Error Image, and Neural Network[C]. Proceedings of the IEEE International Conference on Multimedia & Expo, 2005, pp.1~4.
113. Wang Y, Moulin P. Optimized Feature Extraction for Learning-based Image Steganalysis[J]. IEEE Transactions on Information Forensics and Security, 2007, 2 (1): 31~45.
114. Avcibas I, Kharrazib M, Memon N, Sankur B. Image Steganalysis with Binary Similarity Measures[J]. EURASIP Journal on Applied Signal Processing, 2005, 17: 2749~2757.
115. Chen X C, Wang Y H, Tan T N, Guo L. Blind Image Steganalysis Based on Statistical Analysis of Empirical Matrix[C]. Proceedings of the 18th International Conference on Pattern Recognition, Hong Kong, 2006, pp.1107~1110.
116. Sullivan K, Madhow U, Chandrasekaran S, Manjunath B S. Steganalysis for Markov Cover Data with Applications to Images[J]. IEEE Transactions on Information Forensics and Security, 2006, 1 (2): 275~287.
117. Shi Y Q, Chen C, Chen W. A Markov Process-based Approach to Effective Attacking JPEG Steganography[C]. Proceedings of the 8th Information Hiding Workshop, 2006, 249~264.
118. Lie W N, Lin G S. A Feature-Based Classification Technique for Blind Image Steganalysis[J]. IEEE Transactions on Multimedia, 2005, 7 (6): 1007~1020.
119. Luo X Y, Wang D S, Wang P, Liu F L. A Review on Blind Detection for Image

- Steganography[J]. Signal Processing, 2008, 88 (9): 2138~2157.
120. Lafferty P, Ahmed F. Texture-based Steganalysis: Results for Color Images[C]. Proceedings of the SPIE Mathematics of Data/Image Coding, Compression, and Encryption VII, with Applications, 2004, pp.145~151.
 121. Fridrich J, Goljan M, Du R. Steganalysis Based on JPEG Compatibility[C]. Proceedings of SPIE Multimedia Systems and Applications IV, 2001, Vol. 4518, pp.275~280.
 122. Dumitrescu S, Wu X, Wang Z. Detection of LSB Steganography via Sample Pair Analysis[C]. Proceedings of the 5th International Workshop on Information Hiding, Lecture Notes in Computer Science, 2003, Vol. 2578, pp.355~372.
 123. Ozer H, Avcibas S, Sankur B, Memon N. Steganalysis of Audio Based on Audio Quality Metrics[C]. Proceedings of SPIE, 2003, Vol. 5020, pp. 55~66.
 124. Budhia U, Kundur D. Digital Video Steganalysis Exploiting Collusion Sensitivity[C]. Proceedings of SPIE Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense III, 2004, Vol. 5403, pp. 210~221.
 125. Pankajakshan V, Doerr G, Bora P K. Detection of Motion-Incoherent Components in Video Streams[J]. IEEE Transactions on Information Forensics and Security, 2009, 4 (1): 49~58.
 126. 刘镔, 刘粉林, 杨春芳. 基于帧间共谋的视频隐写分析[J]. 通信学报, 2009, 30 (4): 41~49.
 127. 苏育挺, 王莉莉, 张春田. 一种新型视频信息隐藏分析算法[J]. 东南大学学报, 2007, 37 (1): 164~167.
 128. Su Y, Zhang C, Wang L, Zhang C. A New Video Steganalysis based on Mode Detection[C]. Proceedings of International Conference on Audio, Language and Image Processing, Shanghai, 2008, pp. 1507~1510.
 129. Liu Q, Sung A H, Qiao M. Video Steganalysis Based on the Expanded Markov and Joint Distribution on the Transform Domains-Detecting MSU StegoVideo[C]. Proceedings of the 7th International Conference on Machine Learning and Application, San Diego, California, USA, 2008, pp. 671~674.
 130. Wu J, Zhang R, Chen M, Niu X. Steganalysis of MSU Stego Video Based on Discontinuous Coefficient[C]. Proceedings of the 2nd International Conference on Computer Engineering and Technology, Chengdu, China, 2010, pp. 96~99.
 131. Pankajakshan V, Ho A T S. Improving Video Steganalysis using Temporal Correlation[C]. Proceedings of the 3rd International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Kaohsiung, Taiwan, 2007, pp. 26~28.
 132. Liu B, Liu F, Wang P. Inter-frame Correlation based Compressed Video Steganalysis[C]. Proceedings of International Conference on Congress on Image and Signal Processing, 2008, 42~46.
 133. Su J K, Giro B. On the Imperceptibility and Robustness of Digital Fingerprints[C]. Proceedings of the IEEE International Conference on Multimedia Computing and Systems, 1999, Vol.2, pp.530~535.
 134. Su J K, Eggers J J, Girod B. Optimum Attack on Digital Watermarks and Its Defense[C]. Thirty-Fourth Asilomar Conference on Signals, Systems and Computers, 2000, Vol. 2, pp.1836~1840.
 135. Cox I J, Linnartz J P M G. Some General Methods for Tampering with Watermarks[J]. IEEE Journal On Selected Areas in Communications, 1998, 16 (4): 587~593.
 136. Boeufand J, Stern J P. An Analysis of One of the SDMI Candidates[C]. Proceedings of International Workshop on Information Hiding, Lecture Notes in Computer Science, 2001, Vol. 2137, pp. 395~410.
 137. Zhao H, Wu M, Wung Z J, Liu K J R. Nonlinear Collusion Attacks on Independent Fingerprints for Multimedia[C]. IEEE International Conference on Acoustics, Speech and Signal Processing, 2003, Vol. 5, pp. 664~667.
 138. Wu M, Liu B. Attacks on Digital Watermarks[C]. Proceedings of the Thirty-Third Asilomar

- Conference on Signals, Systems, and Computers, 1999, Vol. 2, pp. 1508~1512.
139. Barnett R, Pearson D E. Frequency Mode LR Attack Operator for Digitally Watermarked Images[J]. Electronics Letters, 1998, 34 (19): 1837~1839.
 140. Petitcolas F A P, Kirovski D. The Blind Pattern Matching Attack on Watermark Systems[C]. IEEE International Conference on Acoustics, Speech and Signal Processing, 2002, Vol.4, pp. 3740~3743.
 141. Langelaar G C, Lagendijk R L, Biemond J. Removing Spatial Spread Spectrum Watermarks by Nonlinear Filtering[C]. Proceedings of EUSIPCO '98, 1998, Vol. 4, pp. 2281~2284.
 142. Bogerr B P, Healy M J R, Tukey J W. The Quefrency Analysis of Time Series for Echoes: Cepstrum, Pseudo-autocovariance, CrossCepstrum, and Saphe Cracking[C]. Proceedings of Symposium on Time Series Analysis, 1963, pp. 209~243.
 143. Chan C K, Cheng L M. An Attack on the Hwang-Chang-Hwang Watermarking Scheme[J]. IEEE Transactions on Consumer Electronics, 2000, 46 (1): 40~43.
 144. Hwang M S, Chang C C, Hwang K F. A Watermarking Technique Based on One-way Hash Functions[J]. IEEE Transactions on Consumer Electronics, 1999, 45 (2): 286~294.
 145. Licks V, Ourique F, Jordan R, Perez-Gonzalez F. The Effect of the Random Jitter Attack on the Bit Error Rate Performance of Spatial Domain Image Watermarking[C]. International Conference on Image Processing, 2003, Vol. 2, pp.455~458.
 146. Linnartz J P M G, van Dijk M. Analysis of the Sensitivity Attack Against Electronic Watermarks in Images[C]. International Workshop on Information Hiding, Portland, OR, USA, Lecture Notes in Computer Science, 1998, Vol. 1525, pp.258~272.
 147. Kalker T, Linnartz J P M G, van Dijk M. Watermark Estimation Through Detector Analysis[C]. Proceedings of the International Conference on Image Processing, 1998, Vol. 1, pp. 425~429.
 148. O'Ruanaidh J J K, Pereira S. A Secure Robust Digital Image Watermark[J]. Proceedings of SPIE Electronic Imaging: Processing, Printing, and Publishing in Color, 1998, Vol.3409, pp.150~163.
 149. Kutter M, Voloshynovskiy S, Herrigel A. The Watermark Copy Attack[C]. Proceedings of SPIE Security and Watermarking of Multimedia Contents II, 2000, Vol. 3971, pp.371~380.
 150. Craver S, Memon N, Yeo B L, Yeung M M. Resolving Rightful Ownerships with Invisible Watermarking Techniques: Limitations, Attacks and Implications[J]. IEEE Journal on Selected Areas in Communications, 1998, 16 (4): 573~586.
 151. Barreto P S L M, Kim H Y, Rijmen V. Toward a Secure Public-Key Blockwise Fragile Authentication Watermarking[C]. Proceedings of the IEEE International Conference on Image Processing, Thessaloniki, Greece, 2001, Vol. 2, pp. 494~497.
 152. Wong P W. A Public Key Watermark for Image Verification and Authentication[C]. Proceedings of the International Conference on Image Processing, Chicago, 1998, Vol. 1, pp. 425~429.
 153. Holliman M, Memon N. Counterfeiting Attacks on Oblivious Block-wise Independent Invisible Watermarking Schemes[J]. IEEE Transactions on Image Processing, 2000, 9 (3): 432~441.
 154. 张沙琦. 数字作品知识产权保护的技术措施研究[D]. 华中师范大学硕士学位论文, 2006.
 155. 赵倩. 数字水印版权保护系统[D]. 首都经济贸易大学硕士学位论文, 2004.
 156. 杨伟, 王飞, 张中, 杨义先, 钮心忻. 伪装式数字化语音保密通信系统[J]. 通信学报, 2004, 25 (2) 75~81.
 157. 杨青春. 视频作品版权追踪系统的设计与实现[D]. 华中科技大学硕士学位论文, 2010.
 158. 杨丽杰. 用于图像认证的安全易损水印算法研究[D]. 燕山大学硕士学位论文, 2006.
 159. 张婧. 基于复制控制的视频数字水印嵌入提取算法的研究[D]. 吉林大学硕士学位论文, 2004.
 160. 孙颖. 视频水印在数字视频广播系统监控中的应用[D]. 山东大学硕士学位论文, 2008.

索引

A

Absolute Moment BTC (AMBTC)
Active attack
Active warder
Adaptive chosen plaintext attack
Advanced Video Coding (AVC)
Alternating Current (AC)
Analysis of Variance (ANOVA)
Annotation
Anonymous communication
Anonymous fingerprinting
Area Under Curve (AUC)
Asymmetric fingerprinting
Asymmetric information hiding
Asymmetric steganography
Authorship mark
Automatic Repeat reQuest (ARQ)
Availability

B

Balanced Incomplete Block Design (BIBD)
Bark Spectral Distortion (BSD)
Benchmarking
Binary Frequency Shift Keying (2FSK)
Binary Phase Shift Keying (BPSK)
Binary Similarity Measures (BSM)
Bit Error Rate (BER)
Bitmap (BMP)
Bit-plane
Bit-Plane Complexity Segmentation (BPCS)
Bit-shifting
Blind and informed detection
Blind information hiding
Blind Pattern Matching (BPM)
Block Truncation Coding (BTC)
Bloom filter
Brute-force attack

C

Calibrated image
Carrier system
Cat map

绝对矩块截断编码 §5.6.2
主动攻击 §2.3.3
主动看守者 §1.4.1
自适应选择明文 §1.2.3
高级视频编码 §1.5.3
交流 §2.5.3
方差分析法 §7.5.3
标注 §1.7.5
匿名通信 §1.2.2
匿名指纹技术 §4.2.3
区域的面积 §7.2.2
非对称指纹技术 §4.2.3
非对称(公钥)信息隐藏 §1.2.2
非对称(公钥)隐写术 §2.2.4
身份标识水印 §9.7.2
自动重复请求 §6.3.2
可用性 §1.1.2

均衡不完全区组设计 §4.9.6
Bark 谱失真 §7.5.2
测试基准 §3.4.3
二元频移键控 §6.3.1
二相相移键控 §3.7.3
二元相似性度量 §7.4.3
误码率 §2.3.2
位图 §2.5.1
位平面 §5.6.2
位平面复杂度分割隐写术 §2.5.2
比特移位 §5.4.3
盲检测与明检测 §3.4.2
盲隐藏 §1.3.2
盲模式匹配攻击 §8.3.2
块截断编码 §2.5.4
布鲁姆过滤器 §6.4.2
强力攻击 §8.2.1

校正图像 §7.3.3
载体系统 §6.2.1
猫映射 §3.5.3

Center of Mass (CoM)	质心	§7.3.3
Cepstrum Distance Measure (CDM)	倒谱距离测度	§7.5.2
Changeable	可交换的	§4.5.3
Chip rate	片率	§2.2.4
Chirp Spread Spectrum (Chirp-SS)	脉冲线性扩频 (切普扩频)	§6.3.1
Chosen plaintext attack	选择明文攻击	§1.2.3
Chosen secret message attack	选择秘密消息攻击	§1.2.3
Cipher block	加密块	§4.6.2
Cipher text	密文	§1.2.2
Ciphertext-only attack	唯密文攻击	§1.2.3
Code obfuscation	代码模糊技术	§9.7.2
Code tamper	代码篡改	§9.7.2
Code watermark	代码水印	§9.7.2
Collusion attack	共谋攻击	§1.7.3
Collusion-secure	共谋安全的	§4.4.1
Collusion-secure code	共谋安全码	§8.3.3
Column or line removal	行列去除	§8.3.3
Combinatorial design	组合设计	§4.9.6
Common Intermediate Format (CIF)	一种视频格式	§3.7.3
Component Object Model (COM)	组件模型	§9.4.4
Computer science	计算机科学	§1.1.2
Confidentiality	保密性	§1.1.2
Consultative Committee for International Telephony and Telegraphy (CCITT)	CCITT 国际组织	§2.3.1
Content Authentication	内容认证	§1.1.1
Context-based Adaptive Lossless Image Codec (CALIC)	基于上下文自适应无损图像压缩	§5.4.1
Controllability	可控性	§1.1.2
Copy attack	复制攻击	§4.3.2
Copy control	复制控制	§1.7.1
Copyright	版权	§1.1.1
Copyright marking	版权标记	§1.4.2
Copyright protection	版权保护	§1.1.1
Cover data	载体数据	§6.2.1
Cover object	载体对象	§1.2.2
Covert channel	隐蔽信道	§1.4.3
Covert storage channel	隐蔽存储信道	§1.4.3
Covert timing channel	隐蔽时间信道	§1.4.3
Crypto data	密码数据	§6.2.1
Cryptographic attack	密码攻击	§8.2.1
Cryptography	密码术	§1.1.2
D		
Data capacity	数据容量	§3.5.1
Data conflict covert channel	数据冲突隐蔽信道	§6.1.3
Data Encryption Standard (DES)	数据加密标准	§1.4.4
Data hiding	数据隐藏	§1.2.2
Data supplier	发行商	§4.6.2
Date watermark	数据水印	§9.7.2
DCT with HAS (DCTwHAS)	基于 HAS 的 DCT	§7.5.4
Delta Modulation (DM)	增量调制	§6.3.1

Denial of Service (DOS)	拒绝服务攻击	§6.3.1
Difference Expansion (DE)	差值扩展	§5.4.3
Difference Pulse Code Modulation (DPCM)	差分脉码调制	§2.5.4
Differential Image Histogram (DIH)	差值图像直方图	§7.3.2
Digital Audio Broadcast (DAB)	数字声音广播	§9.6.1
Digital fingerprint	数字指纹	§1.4.2
Digital fingerprinting	数字指纹技术	§1.2.2
Digital Multimedia Broadcast (DMB)	数字多媒体广播	§9.6.1
Digital Rights Management (DRM)	数字版权管理	§4.1.2
Digital signature	数字签名	§1.1.2
Digital Video Broadcast (DVB)	数字视频广播	§4.1.2
Digital watermark	数字水印	§1.4.2
Digital watermarking	数字水印技术	§1.2.2
Direct Current (DC)	直流	§2.5.3
Direct Sequence Spread Spectrum (DSSS)	直接序列扩频	§2.6.2
Discrete Cosine Transform (DCT)	离散余弦变换	§1.5.3
Discrete Fourier Transform (DFT)	离散傅里叶变换	§2.5.1
Discrete Wavelet Transform (DWT)	离散小波变换	§1.5.3
Distinguished Name (DN)	标识符	§9.3.3
Dither Modulation (DM)	抖动调制	§2.6.2
Dynamic software watermarking	动态软件水印技术	§9.7.2
E		
Echo Hiding	回声隐藏	§2.6.2
Effectiveness	有效性	§3.4.2
Embedded object	嵌入对象	§1.2.2
Embedding algorithm	嵌入算法	§1.2.2
Embedding capacity	嵌入容量	§1.5.4
Embedding key	嵌入密钥	§1.2.2
Embeddor	嵌入者	§1.2.2
Enabling block	使能块	§4.6.2
Entropy encoding	熵编码	§2.5.4
European Broadcasting Union (EBU)	欧洲广播联盟	§3.5.3
Expandable	可扩展的	§5.4.3
Extracting algorithm	提取算法	§1.2.2
Extracting key	提取密钥	§1.2.2
Extractor	提取者	§1.2.2
EzStego	一种调色板隐写术	§2.5.2
F		
F3	一种 JPEG 图像隐写方法	§2.5.4
F4	一种 JPEG 图像隐写方法	§2.5.4
F5	一种 JPEG 图像隐写方法	§2.5.1
Facial Animation Parameter (FAP)	脸部运动参数	§3.7.4
Fact database	事实数据库	§2.4.5
False negative	漏检	§4.2.2
False negative rate	漏检率	§7.6.3
False positive	虚检	§4.2.2
False positive behavior	虚检行为	§3.4.2
False positive rate	虚检率	§7.6.3

Fast Fourier Transform (FFT)	快速傅里叶变换	§1.5.3
Fidelity	保真度	§3.4
Fingerprint	指纹	§1.4.2
Fingerprinting	指纹识别/指纹技术	§4.1.1
Fingerprinting mark	指纹水印	§9.7.2
Fisher Linear Discriminant (FLD)	Fisher 线性分类器	§7.4.1
Fragile information hiding	脆弱信息隐藏	§1.5.4
Fragile watermark	脆弱水印	§1.7.4
Fragile watermarking	脆弱水印技术	§1.7.4
Fragility	脆弱性	§3.5.1
Frameproof (FP)	防陷害码	§4.4.1
Frequency Hopping (FH)	跳频	§6.3.1
Frequency Hopping Spread Spectrum (FHSS)	跳频扩频	§6.3.1
Frequency Mode Laplacian Removal (FMLR)	频率模式拉普拉斯去除	§8.3.2
Frequency Modulation (FM)	调频	§6.3.1
G		
Geometric attack	几何攻击	§8.3.3
Geometric open code	几何隐语	§2.2.3
Geometrical distortion attack	几何失真攻击	§4.3.2
General LSB Lossless Compression	广义最低有效位无损压缩	§5.4.1
General information security	广义信息安全	§1.1.2
Graphics Interchange Format (GIF)	图像互换格式	§2.5.1
H		
H.26X	系列视频编码的国际标准	§1.5.3
Harmless message	无害消息	§6.2.1
Hider	隐藏者	§1.2.2
Hiding capacity	隐藏容量	§1.2.3
Hierarchical encoding	层次编码	§2.5.4
Histogram Characteristic Function (HCF)	直方图特征函数	§7.3.3
Horizontal reflection	水平静像	§8.3.3
Human Audio System (HAS)	人耳听觉系统	§2.6.1
Human Visual System (HVS)	人类视觉系统	§1.5.1
I		
Identifiable Parent Property (IPP)	可确认父元码	§4.4.1
Imperceptibility	不可感知性	§1.2.3
Independent Component Analysis (ICA)	独立分量分析	§2.7.2
Information embedding	信息嵌入	§1.2.2
Information extracting	信息提取	§1.2.2
Information hiding	信息隐藏	§1.1.2
Information security	信息安全	§1.1.1
Information security technology	信息安全技术	§1.1.2
Information theory	信息论	§1.1.2
Integrity	完整性	§1.1.2
Intellectual property right protection	知识产权保护	§1.2.1
Inter-Frame Correlation-based Steganalysis (IFCS)	帧间相关视频隐写分析	§7.6.4
International Federation of Phonographic Industry (IFPI)	国际唱片业协会	§3.6.4
Internet	因特网 (国际互联网)	§1.1.1

Interpretation attack	解释攻击	§4.3.2
Inter-Process Communication (IPC)	进程间通信	§1.4.3
Inverse Discrete Cosine Transform (IDCT)	离散余弦变换逆变换	§2.5.3
Inverse Discrete Fourier Transform (IDFT)	离散傅里叶逆变换	§2.5.3
Irreversible information hiding	不可逆信息隐藏	§1.5.4
Itakura-Siata Distance (ISD)	Itakura-Siata 距离	§7.5.2
J		
Jitter attack	抖动攻击	§4.3.2
Joint Bi-level Image experts Group (JBIG)	联合二值图像专家组	§5.4.1
Joint Photographic Experts Group (JPEG)	联合图像专家组	§1.5.3
JPEG2000	基于小波的图像压缩标准	§2.5.2
JPHide&Seek	一种 JPEG 图像隐写方法	§2.5.1
JSteg	一种 JPEG 图像隐写方法	§2.5.1
Judge	仲裁者	§4.6.2
Just Noticeable Difference (JND)	刚好可察觉误差	§3.3.3
K		
Key bitplane	关键位平面	§5.4.1
Known cover object attack	已知载体对象攻击	§1.2.3
Known plaintext attack	已知明文	§1.2.3
Known secret message attack	已知秘密消息攻击	§1.2.3
I		
Image Quality Metrics (IQMs)	图像质量度量	§7.4.3
Integer DCT	整数余弦变换	§5.5
Integer DWT	整数小波变换	§5.5
Item	商品	§4.6.2
L		
Least Significant Bit (LSB)	最不重要位 (最低有效位)	§1.5.3
Least Two Significant Bits (LTSB, L2SB)	最不重要两位	§2.5.2
Legal attack	合法攻击 (法律攻击)	§8.6.1
Licensing mark	许可证水印	§9.7.2
Linear Predictive Coding (LPC)	线性预测编码	§7.5.2
Linguistic steganography	语义隐写术	§1.4.1
Local Binary Pattern (LBP)	局部二值模式	§7.4.3
Location map	位置地图	§5.4.3
Log Likelihood Ratio (LLR)	对数似然比	§7.5.2
Lossless bit-plane compression	无损位平面压缩	§5.4.1
Lossless information hiding	无损信息隐藏	§5.2.1
Lossless encoding	无损真编码	§2.5.4
Low probability intercept communication	低截获概率通信	§1.2.2
LSB matching	LSB 匹配隐写术	§2.5.2
M		
Marking assumption	嵌入假设	§4.2.2
Marking space	标记空间	§3.3.3
Mean Opinion Score (MOS)	MOS 得分	§2.3.1
Mean Square Error (MSE)	均方误差	§2.3.1

Media bridge	媒体桥	§9.7.3
Media space	媒体空间	§3.3.3
Message layer	消息层	§8.5.2
Meteor-burst communication	流星猝发（余迹散射）通信	§1.4.5
Minimum Shift Keying (MSK)	最小相移键控	§6.3.1
Model Based steganography (MB)	基于模型的隐写术	§2.5.4
Modified BSD (MBSD)	改进的 Bark 谱失真	§7.5.2
Modified Median Edge Detector (MMED)	改进型 MED 预测器	§2.5.2
Modified Pixel Value Differencing (MPVD)	改进像素差隐写术	§2.5.2
Moment Preserving Quantizer (MPQ)	矩保留量化器	§5.6.2
Mosaic attack	马赛克攻击	§4.3.2
Moving Picture Experts Group (MPEG)	运动图像专家组	§1.5.3
MP3 (MPEG-1 or MPEG-2 Audio Layer III)	MP3	§1.5.3
Mp3Stego	一种针对 MP3 隐写软件	§2.6.4
MSU StegoVideo	一种视频隐写软件	§7.6.4
Multilevel secure system	多级安全系统	§1.4.3
Multiple Base Notational System (MBNS)	混合进制系统隐写术	§2.5.2
N		
Narrow information security	狭义信息安全	§1.1.2
Networking era	网络时代	§1.1.1
Non-blind information hiding	非盲隐藏	§1.3.2
Non-framing	不可诬陷性	§4.6.2
Non-repudiation	不可否认（抵赖）性	§1.1.2
Normalized Correlation (NC)	归一化相关	§2.3.2
Normalized Expected Cost (NEC)	正规化期望代价	§7.2.3
Null cipher	虚字密码	§1.4.1
O		
Oblivious watermarking	盲提取水印方案	§3.7.3
Object Linking and Embedding (OLE)	对象链接和嵌入	§9.4.4
Objective Difference Grade (ODG)	主观差别等级	§2.3.1
Onion Router (OR)	洋葱路由	§6.4.2
Ontological semantics	本体语义	§2.4.5
Open code	隐语	§1.4.1
Open Systems Interconnection (OSI)	开放系统互联	§6.1.3
Optimal hyperplane	最优分类面	§7.4.2
Original carrier	原始载体	§1.3.2
OutGuess	一种 JPEG 图像隐写方法	§2.5.1
Overt channel	公开信道	§1.4.4
Overt receiver	公开收方	§1.4.4
P		
Pairs Of Value (POV)	值对	§7.3.4
Particle Swam Optimization (PSO)	粒子群优化算法	§7.5.3
Passive attack	被动攻击	§2.3.4
Passive attacker	被动攻击者	§2.2.4
Passive warder	被动看守者	§1.4.1
Payload	容量	§3.4.2
Peak point	峰点	§5.4.4

Peak Signal to Noise Ratio (PSNR)	峰值信噪比	§2.3.1
Perceptual redundancy	感知冗余	§1.3.4
Perceptual space	感知空间	§1.3.4
Perceptual system	感知系统	§1.3.4
Personal key	个人密钥	§4.6.2
Phase coding	相位编码	§2.6.3
Phase dispersion	相位离散	§2.6.3
Pirate	盗版者	§1.4.2
Pirate user	非法用户	§4.6.2
Pitch-preserving scaling	音调保持缩放	§8.3.3
Pixel Value Differencing (PVD)	像素差隐写术	§2.5.2
Plaintext	明文	§1.4.1
Port Hopping (PH)	跳端口	§6.3.1
Prediction Error Frame (PEF)	预测残差帧	§7.6.4
Predictive Coding Based Steganography (PCBS)	预测编码隐写术	§2.5.2
Presentation Attack	表达攻击	§8.2.1
Principal Component Analysis (PCA)	主成分分析法	§7.5.3
Private key	私钥	§4.6.2
Private watermark	私有水印	§8.1.2
Private watermarking	私有水印处理	§8.1.2
Private watermarking algorithm	私有水印算法	§8.1.2
Probability Cost Function (PCF)	概率花费函数	§7.2.3
Progressive encoding	递增式编码	§2.5.4
Projection Histogram (PH)	投影直方图	§7.4.3
Protocol attack	协议攻击	§4.3.2
Pseudo-noise sequence	PN 序列	§2.6.2
Public-key cryptography	公钥密码术	§8.1.2
Public Key Infrastructure (PKI)	公钥基础设施	§9.1
Public watermark	公有水印	§8.1.2
Public watermarking	公有水印处理	§8.1.2
Public watermarking algorithm	公有水印算法	§8.1.2
Pulse Code Modulation (PCM)	脉冲编码调制	§2.6.2
Pure steganography	纯隐写术	§2.2.4

Q

Quadrature Mirror Filter (QMF)	正交镜像滤波器	§7.4.3
Quadrature Phase Shift Keying (QPSK)	四相相移键控	§6.3.1
Qualified Significant Wavelet Tree (QSWT)	重要树	§3.5.3
Quantization Index Modulation (QIM)	量化索引调制	§2.6.2

R

Receiver Operating Characteristic (ROC)	接收操作特性	§7.2.2
Recipient anonymity	接收方匿名	§6.4.1
Recording Industry Association of America (RIAA)	美国唱片协会	§3.6.4
Recording space	记录空间	§1.3.4
Recording system	记录系统	§1.3.4
Regular groups and Singular groups (RS)	RS 检测法	§7.3.2
Regularity-Singularity (R-S)	R-S 无损信息隐藏算法	§5.4.1
Removal attack	去除攻击	§4.3.2
Reserve engineering	逆向工程	§9.7.2

Reversible data hiding	可逆数据隐藏	§5.2.1
Reversible information hiding	可逆信息隐藏	§1.5.4
Robust information hiding	鲁棒信息隐藏	§1.5.4
Robust watermark	鲁棒水印	§1.7.2
Robust watermarking	鲁棒水印技术	§1.7.2
Robustness	鲁棒性 (稳健性)	§1.2.3
Robustness attack	鲁棒性攻击	§8.2.1
RSA (Rivest-Shamir-Adleman)	一种非对称加密算法	§2.1.2
Run length	游程长度 (行程长度)	§7.3.3
Run-Length Encoding (RLE)	行程长度编码	§2.5.4
S		
Sample Pairs Analysis (SPA)	抽样对分析法	§7.5.1
Sample removal	样值去除	§8.3.3
Scale Factor Band (SFB)	尺度因子带	§3.6.7
Scrambling attack	拼凑攻击	§8.3.2
Secret communication	保密通信	§1.2.1
Secret key	密钥	§3.4.2
Secret information (message)	秘密消息 (信息)	§1.1.2
Secure Digital Music Initiative (SDMI)	安全数字音乐倡导者联盟	§3.6.4
Secure FP (SFP)	安全防陷害码	§4.4.1
Security	安全性	§1.1.2
Security mechanism	安全机制	§1.1.2
Segmental Signal-to-Noise Ratio (SNRseg)	分段信噪比	§2.3.1
Self-recoverability	自恢复性	§1.2.3
Semagram	符号码	§1.4.1
Semi-fragile information hiding	半脆弱信息隐藏	§1.5.4
Sender anonymity	发送方匿名	§6.4.1
Sensitivity analysis attack	敏感分析攻击	§8.3.3
Sequential encoding	顺序式编码	§2.5.4
Sequential Floating Search (SFS)	SFS 方法	§7.5.3
Session block	会话数据块	§4.6.2
Session key	会话密钥	§4.6.2
Shearing	剪切	§8.3.3
Side information	边信息 (附加信息)	§1.3.3
Signal-to-Noise Ratio (SNR)	信噪比	§3.6.3
Single copy attack	单一复制攻击	§4.2.2
Static software watermarking	静态软件水印技术	§9.7.2
Software piracy	软件盗版	§9.7.2
Software tamper-proofing	软件防篡改技术	§9.7.2
Software watermarking	软件水印技术	§9.7.2
Sound Pressure Level (SPL)	声压级	§2.6.1
Spectral Phase Distortion (SPD)	频谱相位失真	§7.5.2
Spectral Phase Magnitude Distortion (SPMD)	频谱相位幅度失真	§7.5.2
Spread spectrum	扩频	§7.4.3
Spread spectrum communication	扩展频谱通信 (扩频通信)	§1.4.5
Spread Spectrum Image Steganography (SSIS)	扩频图像隐写术	§2.5.2
Star property	*-特性	§6.1.1
Statistical Learning Theory (SLT)	统计学习理论	§7.4.2
Steganography	隐写术 (隐藏技术)	§1.2.2

Stego capacity	隐写容量	§1.4.1
Stego carrier	含密载体	§1.3.2
Stego data	隐写数据	§6.2.1
Stego key	隐藏密钥 (隐写密钥)	§1.2.2
Stego object	伪装对象 (隐写对象)	§1.2.2
Stego object-only attack	唯伪装对象攻击	§1.2.3
Stegoanalysis	隐藏分析 (伪装分析)	§1.2.2
Stegoanalyst	隐藏分析者 (伪装分析者)	§1.2.2
StegoDos	一种空域隐写软件	§2.5.2
STools	一种空域隐写软件	§2.5.2
Subliminal channel	阈下信道 (潜信道)	§1.4.4
Subliminal information (message)	阈下信息 (消息)	§1.4.4
Subliminal receiver	阈下收方	§1.4.4
Subliminal sender	阈下发方	§6.2.1
Sub-pixel Accuracy Motion Compensation Prediction (SPAMCP)	亚像素运动补偿	§7.6.4
Subtractive Pixel Adjacency Model (SPAM)	差值像素邻域模型	§7.3.3
Support Vector (SV)	支持向量	§7.4.2
Support Vector Machine (SVM)	支持向量机	§7.4.1
Symmetric fingerprinting	对称指纹技术	§4.2.3
Symmetric information hiding	对称 (私钥) 信息隐藏	§1.2.2
Symmetric steganography	对称 (私钥) 隐写术	§2.2.4
Synchronization attack	同步攻击	§4.3.2
System attack	系统攻击	§8.6.12
T		
Technical steganography	技术隐写术	§1.4.1
Temporal Frames Averaging (TFA)	时域帧平均	§7.6.4
Text Meaning Representation (TMR)	文本含义描述	§2.4.5
Time Hopping Spread Spectrum (THSS)	跳时扩频	§6.3.1
Time stamp	时间戳	§8.4.6
Tracibility (TA)	可追踪码	§4.4.1
Traitor	叛逆者	§1.4.2
Traitors	共谋者	§4.6.2
Traitor tracing	叛逆者追踪 (盗版跟踪)	§1.7.1
Transparency	透明性	§1.2.3
Transport layer	传输层	§8.5.2
True Positive Rate (TPR)	击中率	§7.6.3
Trusted Third Part (TTP)	可信第三方	§4.6.2
U		
Unambiguous	确定性	§3.5.1
Unauthorized detection	非授权检测	§8.2.1
Unauthorized embedding	非授权嵌入	§8.2.1
Unauthorized removal	非授权去除	§8.2.1
Undetectability	不可检测性	§1.2.3
Unlinkability of sender and recipient	收发双方无关联	§6.4.1
Unpredictable Randomness	不可预测的随机性	§3.5.3
User	合法用户	§4.6.2
User Interface (UI)	用户接口	§8.6.2

V

Validation mark	验证水印	§9.7.2
Variable Length Coding (VLC)	可变长编码	§2.7.3
Variable Length Integer (VLI)	可变长整数	§5.6.3
Vector Quantization (VQ)	矢量量化	§2.5.4
Vertical Blanking Interval (VBI)	垂直空白间隔	§3.4.1
Very High Frequency (VHF)	甚高频	§6.3.2
Video Object (VO)	视频对象	§2.7.3
Video on Demand (VOD)	视频点播	§4.1.2
Visible Watermarking	可见水印技术	§1.2.3

W

Warden	看守者	§1.3.1
Watermark	水印	§1.4.2
Web camera	网络摄像机	§9.7.3
Web-crawling detector	网页过滤检测器	§8.3.3
Weighted Spectral Slope Distance (WSSD)	加权频谱斜率距离	§7.5.2
White noise storm	一种空域隐写软件	§2.5.2
Wiener filtering	维纳滤波	§8.3.2

Z

Zero point	零点	§5.4.4
Zerotree	零树	§3.5.3